

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2005, 08:17:51 ; Search time 182 Seconds

(without alignments)

163.190 Million cell updates/sec

Title: US-09-980-672-2

Perfect score: 327

Sequence: 1 SHLVKCAEKTFCVNGGECFM.....TGDRCQNVYMASFYKABELY 58

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 2430

Minimum DB seq length: 58

Maximum DB seq length: 58

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: uniprot\_03.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	160	48.9	58	Q86WJ0	homo sapien
2	43.5	13.3	58	MTL_HOMAM	P29499 homarus ame
3	43.5	13.3	58	Q95P38	homarus ame
4	40	12.2	58	Q69BD0	campylobact
5	40	12.2	58	Q69B10	campylobact
6	39.5	12.1	58	Q7S6X3	neurospora
7	39	11.9	58	MTL_SCYSE	scylla serr
8	39	11.9	58	MT2_CALSI	callinectes
9	39	11.9	58	Q8ABR7	bacteroides
10	37.5	11.5	58	Q7QP67	giardia lam
11	37.5	11.5	58	Q7VBA7	prochloroco
12	37	11.3	58	MT_POTPO	potamon pot
13	37	11.3	58	Q7RQ13	plasmodium
14	37	11.3	58	Q9U623	pacifastacu
15	36.5	11.2	58	MT_CARMA	carcinus ma
16	36.5	11.2	58	Q86M28	carcinus ma
17	36	11.0	58	MT_ASHFL	astacus flu
18	35.5	10.9	58	Q5U91	scylla serr
19	35	10.7	58	Q75J19	dictyosteli
20	35	10.6	58	Q9QD80	human immun
21	34.5	10.6	58	TX20_PHORI	phoneutria
22	34.5	10.6	58	TX34_PHORI	phoneutria
23	34	10.4	58	CTC7_HUMAN	homo sapien
24	34	10.4	58	Q6K4G1	oryza sativ
25	33.5	10.2	58	Q8GU49	posidonina o
26	33.5	10.2	58	Q8H494	oryza sativ
27	33.5	10.2	58	Q9QA93	gb virus c/
28	33	10.1	58	Q6A1C0	vigna subte
29	33	10.1	58	Q8XPP2	ralstonia s
30	33	10.1	58	Q97JP4	clostridium
31	32.5	9.9	58	Y603_PYRAB	pyrococcus

32	32.5	9.9	58	Q8IDR2	plasmodium
33	32.5	9.9	58	Q7RPL4	plasmodium
34	32.5	9.9	58	Q7LZB6	anas platyr
35	32	9.8	58	Q8TEX8	homo sapien
36	32	9.8	58	Q9BZ93	homo sapien
37	32	9.8	58	Q95U92	eriocheir s
38	32	9.8	58	Q95U93	portunus pe
39	32	9.8	58	Q64UX2	bacteroides
40	32	9.8	58	Q92TU4	rhizobium m
41	32	9.8	58	Q97DF0	clostridium
42	32	9.8	58	Q73PH3	treponema d
43	32	9.8	58	Q91GI9	epiphyas po
44	31.5	9.6	58	Q8XL34	clostridium
45	31.5	9.6	58	Q81S98	clostridium
46	31.5	9.6	58	Q83GS6	bacillus an
47	31	9.5	58	Q7RKL5	tropheryma
48	31	9.5	58	Q6ALC4	plasmodium
49	31	9.5	58	Q51596	canavalia b
50	31	9.5	58	Q83033	pelobacter
51	31	9.5	58	Q83033	banana bunc
52	31	9.5	58	Q9QA82	gb virus c/
53	30.5	9.3	58	Q9WJ7	gb virus c/
54	30.5	9.3	58	T223_PHONI	phoneutria
55	30.5	9.3	58	TX23_PHOKE	phoneutria
56	30.5	9.3	58	Q8S1V2	oryza sativ
57	30	9.2	58	Q7P4T3	fusobacteri
58	30	9.2	58	FER6_METJA	methanococc
59	30	9.2	58	TAT_FV1B5	human immun
60	30	9.2	58	Q95MF3	sus scrofa
61	30	9.2	58	Q6A1B9	vigna ungui
62	30	9.2	58	Q48577	leptosira
63	29.5	9.0	58	Q92H84	rickettsia
64	29.5	9.0	58	Q68DW6	homo sapien
65	29.5	9.0	58	Q7WGO	anemonia su
66	29.5	9.0	58	Q64ZC7	bacteroides
67	29	8.9	58	Q83379	treponema p
68	29	8.9	58	Q96X19	sulfolobus
69	29	8.9	58	Q9BYX8	homo sapien
70	29	8.9	58	Q95N07	bos indicus
71	29	8.9	58	Q8MAW6	cuscuta eur
72	29	8.9	58	Q853K9	mycobacteri
73	29	8.9	58	Q9AK80	streptomyce
74	28.5	8.7	58	Q9EF64	human immun
75	28.5	8.7	58	Q96ZK5	sulfolobus
76	28.5	8.7	58	Q8XSD9	ralstonia s
77	28	8.6	58	Q825D1	streptomyce
78	28	8.6	58	Q6KZV4	parabuthus
79	28	8.6	58	Q9N6X0	picrophilus
80	28	8.6	58	Q6TK31	caenorhabdi
81	28	8.6	58	Q6TK32	praecitrull
82	28	8.6	58	Q6TK33	sechium edu
83	28	8.6	58	Q6TK34	sicyos angu
84	28	8.6	58	Q6TK35	trichosanthe
85	28	8.6	58	Q6TK36	luffa grave
86	28	8.6	58	Q6TK37	luffa echin
87	28	8.6	58	Q6TK38	cucurbita p
88	28	8.6	58	Q6TK39	benincasa h
89	28	8.6	58	Q6TK40	luffa quinq
90	28	8.6	58	Q6TK41	marah orega
91	28	8.6	58	Q6TK42	echinocysti
92	28	8.6	58	Q6TK43	lagenaria l
93	28	8.6	58	Q6TK44	citrullus l
94	28	8.6	58	Q6TK45	citrullus c
95	28	8.6	58	Q6TK46	acanthosicy
96	28	8.6	58	Q6TK47	bryonia dio
97	28	8.6	58	Q6TK48	coccinia pa
98	28	8.6	58	Q6TK49	diplocyclos
99	28	8.6	58	Q6TK50	eballium e
100	28	8.6	58	Q6SKD3	arthrobacte

ALIGNMENTS

RESULT 1  
 Q86WJO PRELIMINARY; PRT; 58 AA.  
 AC Q86WJO  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Neuregulin 1 isoform 4 (Fragment).  
 GN Name=NRG1;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. PubMed=14632199;  
 RX MEDLINE=22993650; PubMed=14632199;  
 RA Stove C., Stove V., Derycke L., Van Marck V., Mareel M., Bracke M.;  
 RT "The heregulin/human epidermal growth factor receptor as a new growth  
 factor system in melanoma with multiple ways of deregulation.";  
 RL J. Invest. Dermatol. 121:802-812(2003).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Stove C.P., Stove V.V., Mareel M.M., Bracke M.B.;  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: Contains 1 EGF-like domain.  
 DR EMBL; AY207002; AAO49724.1; -  
 DR HSP; Q12780; 1HAE.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR006210; IEGF.  
 DR Pfam; PF00008; EGF; 1.  
 DR SMART; SM00181; EGF; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS50026; EGF\_3; 1.  
 KW EGF-like domain.  
 FT NON TER 1  
 SQ SEQUENCE 58 AA; 6299 MW; 48F31F8CE4E5A1FB CRC64;  
 Query Match 48.9%; Score 160; DB 2; Length 58;  
 Best Local Similarity 80.6%; Pred. No. 1.2e-11;  
 Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 12 CVNGCECFWVLDLSPNRYLCKPNEFTGDRCONV 47  
 DB 1 CVNGCECFWVLDLSPNRYLCKPNEFTGDRCTENV 36  
 RESULT 2  
 MT1\_HOMAM  
 ID MT1\_HOMAM STANDARD; PRT; 58 AA.  
 AC P29499;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Metallothionein-1 (CuMT-1).  
 OS Homarus americanus (American lobster).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;  
 OC Nephropoidea; Nephropidae; Homarus.  
 OX NCBI\_TaxID=6706;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Hepatopancreas;  
 RX MEDLINE=89215793; PubMed=2709004; DOI=10.1016/0162-0134(89)84018-8;  
 RA Brower M., Winge D.R., Gray W.R.;  
 RT "Structural and functional diversity of copper-metallothioneins from  
 the American lobster Homarus americanus.";  
 RL J. Inorg. Biochem. 35:289-303(1989).  
 RN [2]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE=94318629; PubMed=8043573;

RA Zhu Z., Derosé E.F., Mullen G.P., Petering D.H., Shaw C.F. III;  
 RT "Sequential proton resonance assignments and metal cluster topology of  
 lobster metallothionein-1";  
 RL Biochemistry 33:8858-8865(1994).  
 CC -!- FUNCTION: Metallothioneins have a high content of cysteine  
 residues that bind various heavy metals. The different forms of  
 lobster metallothioneins may have different biological functions.  
 CC Class I MTS in marine crustacea are involved in the sequestration  
 of elevated levels of heavy-metal ions. Binds 6 metal ions. Known  
 to bind cadmium.  
 CC -!- SIMILARITY: Belongs to the metallothionein superfamily. Family 3.  
 DR PIR; A37039; A37039.  
 DR PDB; 1J5L; NMR; A=1-30.  
 DR PDB; 1J5M; NMR; A=1-28.  
 DR InterPro; IPR003019; Metallothion\_3.  
 DR InterPro; IPR002045; Metallothion\_3.  
 DR Pfam; PF00131; Metallothio; 1.  
 DR PRINTS; PR00858; MTRCUSTACEAN.  
 DR 3D-structure; Cadmium; Copper; Direct protein sequencing;  
 KW Metal-binding; Metal-thiolate cluster.  
 FT DOMAIN 1 28  
 FT DOMAIN 29 58  
 FT METAL 4 4  
 FT METAL 5 5  
 FT METAL 9 9  
 FT METAL 11 11  
 FT METAL 16 16  
 FT METAL 20 20  
 FT METAL 22 22  
 FT METAL 25 25  
 FT METAL 27 27  
 FT METAL 30 30  
 FT METAL 33 33  
 FT METAL 37 37  
 FT METAL 39 39  
 FT METAL 45 45  
 FT METAL 49 49  
 FT METAL 53 53  
 FT METAL 55 55  
 FT METAL 56 56  
 SQ SEQUENCE 58 AA; 5975 MW; 176BAAF60A32F96 CRC64;  
 Query Match 13.3%; Score 43.5; DB 1; Length 58;  
 Best Local Similarity 25.0%; Pred. No. 1.1e+03;  
 Matches 12; Conservative 6; Mismatches 13; Indels 17; Gaps 3;  
 QY 6 CAEKTFCVNGG-----ECFVWVLDLSPNRYLCKPNEFTGDR 43  
 DB 5 CKDKCECAEGGCKTGCKTSCRCAPCEKCTSG-----CKCPSK--DEC 45  
 RESULT 3  
 Q95P38 PRELIMINARY; PRT; 58 AA.  
 ID Q95P38  
 AC Q95P38;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Metallothionein.  
 GN Name=mtH;  
 OS Homarus americanus (American lobster).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;  
 OC Nephropoidea; Nephropidae; Homarus.  
 OX NCBI\_TaxID=6706;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Hepatopancreas;  
 RA Valls M., Bofill R., Gonzalez-Duarte R., Gonzalez-Duarte P.,  
 RA Capdevila M., Arrian S.;  
 RT "A new insight into metallothionein MT classification and evolution.  
 The in vivo and in vitro metal binding features of Homarus americanus  
 recombinant MT.";

```

RL J. Biol. Chem. 276:32385-32843(2001).
RP [2]
RW SEQUENCE FROM N.A.
RA Valls M., Domenech J., Gonzalez-Duarte R., Atrian S.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ401298; CAC42504.1; -.
DR EMBL; AJ251112; CAC80859.1; -.
DR HSP; P29499; IJ5L.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR InterPro; IPR003019; Metallothion.
DR Pfam; PF00131; Metallothio; 1.
DR PRINTS; PR00856; MTRUSTACEAN.
SQ SEQUENCE 58 AA; 6024 MW; 5777E2766E629641 CRC64;

Query Match 13.3%; Score 43.5; DB 2; Length 58;
Best Local Similarity 25.0%; Pred. No. 1.1e+03;
Matches 12; Conservative 6; Mismatches 13; Indels 17; Gaps 3;

QY 6 CAEKTFCVNGG-----ECFMVKDLSNPSRYLCKPCNEFTGDCR 43
Db 6 CKDKCEAEGGKGTGCKTSCRCAPECRTSG----CKCPSK---DEC 46

RESULT 4
Q69BD0 PRELIMINARY; PRT; 58 AA.
AC Q69BD0;
DT 25-OCT-2004 (TREMBlrel. 28, Created)
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
DE Cpp6.
OS Campylobacter jejuni.
OG Plasmid pTet.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=81-176;
RA Batchelor R.A., Pearson B.M., Friis L.M., Guerry P., Wells J.M.;
RT "Comparative DNA sequence analyses of two conjugative tetracycline
resistance plasmids from Campylobacter sp.";
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY394561; AAR29540.1; -.
KW Plasmid.
SQ SEQUENCE 58 AA; 7035 MW; 2ED9689C6BF1AB5E CRC64;

Query Match 12.2%; Score 40; DB 2; Length 58;
Best Local Similarity 25.0%; Pred. No. 2.8e+03;
Matches 8; Conservative 5; Mismatches 9; Indels 10; Gaps 1;

QY 5 KCAEKTFCVNGGECFMVKDLSNPSRYLCKPCN 36
Db 25 KCGEKIYV-----EFQPNKFIKPCN 46

RESULT 5
Q69B10 PRELIMINARY; PRT; 58 AA.
AC Q69B10;
DT 25-OCT-2004 (TREMBlrel. 28, Created)
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
DE Cpp6.
OS Campylobacter coli.
OG Plasmid pCC31.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=195;
RN [1]
RP SEQUENCE FROM N.A.
RA Batchelor R.A., Pearson B.M., Friis L.M., Guerry P., Wells J.M.;

"Comparative DNA sequence analyses of two conjugative tetracycline
resistance plasmids from Campylobacter sp.";
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY394560; AAR29490.1; -.
KW Plasmid.
SQ SEQUENCE 58 AA; 7035 MW; 2ED9689C6BF1AB5E CRC64;

Query Match 12.2%; Score 40; DB 2; Length 58;
Best Local Similarity 25.0%; Pred. No. 2.8e+03;
Matches 8; Conservative 5; Mismatches 9; Indels 10; Gaps 1;

QY 5 KCAEKTFCVNGGECFMVKDLSNPSRYLCKPCN 36
Db 25 KCGEKIYV-----EFQPNKFIKPCN 46

RESULT 6
Q7S6X3 PRELIMINARY; PRT; 58 AA.
AC Q7S6X3;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Predicted protein.
GN Name=NCU05590.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamysellis M., Mauceli E., Bielke C., Rudd S., Frisman D.,
RA Krystofova S., Rasmussen C., Metznerberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmani S.A.,
RA Desouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Navig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nubbaum C., Birren B.;
RL "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RT Nature 0:0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABX01000321; EAA31316.1; -.
SQ SEQUENCE 58 AA; 5863 MW; 5199E0AA0DBCDAF3 CRC64;

Query Match 12.1%; Score 39.5; DB 2; Length 58;
Best Local Similarity 29.4%; Pred. No. 3.2e+03;
Matches 10; Conservative 6; Mismatches 13; Indels 5; Gaps 1;

QY 4 VKCAEKTFCVNGGECFMVKDLSNPSRYLCKPCNE 37
Db 28 IPCAAATGVNGGD-----GIAQSEYVWHPNQ 56

RESULT 7
MTL_SCYSE
ID_MTL_SCYSE STANDARD; PRT; 58 AA.
AC P02805;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Metallothionein-I (MT-I).
OS Scylla serrata (Mud crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;

```



```
Query Match 11.9%; Score 39; DB 2; Length 58;
Best Local Similarity 42.3%; Pred. No. 3.6e+03;
Matches 11; Conservative 2; Mismatches 11; Indels 2; Gaps 1;

Qy 17 ECFVWKDLSNPSRYLCKCPNEFTGDR 42
Db 31 ELVIVNDAS--PKYLDLSVNEFDNDR 54

RESULT 10
Q7QP67 PRELIMINARY; PRT; 58 AA.
AC Q7QP67;
DT 01-MAR-2004 (TRENBLrel. 26, Created)
DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)
DE GLP 563 14950 14774.
OS Giardia_lambila ATCC 50803.
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=184922;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WB C6;
RA Morrison H.G., McArthur A.G., Adam R.D., Aley S.B., Gillin F.D.,
RA Olsen G.J., Sogin M.L.;
RT "Draft sequence of the Giardia lamblia genome.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AAC801000177; EAA36803.1; -.
SQ SEQUENCE 58 AA; 6722 MW; CB93E7606F8EC118 CRC64;

Query Match 11.5%; Score 37.5; DB 2; Length 58;
Best Local Similarity 31.0%; Pred. No. 5.5e+03;
Matches 9; Conservative 4; Mismatches 9; Indels 7; Gaps 2;

Qy 31 LCKC----PNEFTGDRQNY---VNASFY 52
Db 7 ICCLQLFLDEEVLSSKCSAYAVKVLAVFY 35

RESULT 11
Q7VBA7 PRELIMINARY; PRT; 58 AA.
AC Q7VBA7;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Predicted protein.
GN OrderedLocusNames=Pro1190;
OS Prochlorococcus marinus.
OC Bacteria; Cyanobacteria; Prochlorales; Prochlorococcaceae;
OX NCBI_TaxID=1219;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SARG / CNCP 1375 / SS120;
RX MEDLINE=22810154; PubMed=12917486; DOI=10.1073/pnas.1733211100;
RA Duressne A., Salanoubat M., Partensky F., Artiguenave F., Axmann I.M.,
RA Barbe V., Duprat S., Galperin M.Y., Koonin E.V., Le Gall F.,
RA Makarova K.S., Ostrowski M., Oztas S., Robert C., Rogozin I.B.,
RA Scanlan D.J., Tandeau de Marsac N., Weissenbach J., Wincker P.,
RA Wolf Y.I., Hesse W.R.;
RT "Genome sequence of the cyanobacterium Prochlorococcus marinus SS120,
RT a nearly minimal oxyphototrophic genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:10020-10025(2003).
DR EMBL; AE017164; AAQ0235.1; -.
KW Complete proteome.
SQ SEQUENCE 58 AA; 6722 MW; 2BC7F5760F14ACFD CRC64;

Query Match 11.5%; Score 37.5; DB 2; Length 58;
Best Local Similarity 28.1%; Pred. No. 5.5e+03;
```

```
Matches 16; Conservative 11; Mismatches 19; Indels 11; Gaps 4;

Qy 1 SHLVKCAEKTFCVNGGE-CFVWKDLSNPSRYLCKCPNEFTGDRCONYVNASFYKABE 56
Db 9 SELNQICEDAF-VNVKCAEMLQEK-----KCSNQVVIEMLRN--VADFYLSEQE 55

RESULT 12
MT_POTPO STANDARD; PRT; 58 AA.
ID MT_POTPO
AC P55952;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Metallothionein (MT).
OS Potamon potamios.
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Potamoidea; Potamidae; Potamon.
OX NCBI_TaxID=59185;
RN [1]
RP SEQUENCE, AND MASS SPECTROMETRY.
RC TISSUE=Midgut;
RX MEDLINE=97079279; PubMed=8921011;
RA Pedersen S.N., Pedersen K.L., Hoejrup P., Depledge M.H., Knudsen J.;
RT "Primary structures of decapod crustacean metallothioneins with
RT special emphasis on freshwater and semi-terrestrial species.";
RL Biochem. J. 319:999-1003(1996).
CC -!- FUNCTION: Metallothioneins have a high content of cysteine
CC residues that bind various heavy metals. Class I MTS in crustacea
CC are involved in the sequestration of elevated levels of heavy-
CC metal ions.
CC -!- INDUCTION: By cadmium.
CC -!- MASS SPECTROMETRY: MW=6156.8; METHOD=MALDI; RANGE=1-58;
CC NOTE=Ref.1.
CC -!- SIMILARITY: Belongs to the metallothionein superfamily. Family 3.
DR HSSP; P29499; 135L.
DR InterPro; IPR003019; Metalthion_
DR InterPro; IPR002045; Metalthion_3.
DR InterPro; IPR008197; WAP.
DR Pfam; PF00131; Metallothio; 1.
DR PRINTS; PR00003; 4DISULPHCORE.
DR PRINTS; PR00858; MTCRUSTACEAN.
KW Cadmium; Direct protein sequencing; Metal-binding;
KW Metal-thiolate cluster.
FT DOMAIN 1 29 Beta.
FT DOMAIN 30 58 Alpha.
FT METAL 4 4 Cluster B.
FT METAL 5 5 Cluster B.
FT METAL 10 10 Cluster B.
FT METAL 12 12 Cluster B.
FT METAL 17 17 Cluster B.
FT METAL 21 21 Cluster B.
FT METAL 23 23 Cluster B.
FT METAL 26 26 Cluster B.
FT METAL 28 28 Cluster B.
FT METAL 31 31 Cluster A.
FT METAL 34 34 Cluster A.
FT METAL 38 38 Cluster A.
FT METAL 40 40 Cluster A.
FT METAL 46 46 Cluster A.
FT METAL 50 50 Cluster A.
FT METAL 54 54 Cluster A.
FT METAL 56 56 Cluster A.
FT METAL 57 57 Cluster A.
SQ SEQUENCE 58 AA; 6157 MW; DCB71F9AB4DF779C CRC64;

Query Match 11.3%; Score 37; DB 1; Length 58;
Best Local Similarity 30.0%; Pred. No. 6.3e+03;
Matches 9; Conservative 4; Mismatches 11; Indels 6; Gaps 2;

Qy 5 KCAEKTFCVNGGECFVWKDLSNPSRYLCKC 34
|| : : : || : : ||
```

```
Db 33 KCTSECKSKKEC--AKNCTKP-----CSC 56

RESULT 13
Q7RQI3 PRELIMINARY; PRT; 58 AA.
AC Q7RQI3;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=PY01114;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteza M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AABL01000295; EAA20367.1; -.
KW Hypothetical protein.
SQ SEQUENCE 58 AA; 6388 MW; B9C0C8EFB8021660 CRC64;

Query Match 11.3%; Score 37; DB 2; Length 58;
Best Local Similarity 28.6%; Pred. No. 6.3e+03;
Matches 10; Conservative 5; Mismatches 10; Indels 10; Gaps 2;

Oy 13 VNGGECFVWKDLSNPSRYLCKPNEFTGDRQNYV 47
Db 10 INGHTLIIVENCEN-----CEN---GENCENVV 34

RESULT 14
O9U623 PRELIMINARY; PRT; 58 AA.
AC Q9U623;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Metallothionein.
OS Pacifastacus leniusculus (Signal crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Astacoidae; Astacidae; Pacifastacidae.
OX NCBI_TaxID=6720;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thoracic ganglia;
RA Skorupek P., Dawbarn D.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF199482; AAUF07215.1; -.
DR HSSP; P29499; 1J5L.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR InterPro; IPR003019; Metallothion_
DR InterPro; IPR002045; Metallothion_3.
DR Pfam; PF00131; Metallothio; 1.
DR PRINTS; PR00858; MTCRUSTACEAN.

SQ SEQUENCE 58 AA; 6008 MW; 576635ACFB0BES100 CRC64;

Query Match 11.3%; Score 37; DB 2; Length 58;
Best Local Similarity 33.3%; Pred. No. 6.3e+03;
Matches 10; Conservative 2; Mismatches 12; Indels 6; Gaps 2;

Oy 5 KCAEKTFCVNGGECFVWKDLSNPSRYLCKC 34
Db 33 KCTSGCKCPKSEEC--ARTCSKP-----CRC 56

RESULT 15
MT_CARMA STANDARD; PRT; 58 AA.
AC P55948;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Metallothionein (MT).
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RC TISSUE=Midgut;
RX MEDLINE=9415337; PubMed=8110201;
RA Pedersen K.L., Pedersen S.N., Hoesrup P., Andersen J.S.,
RA Roepstorff P., Knudsen J., Depledge M.H.;
RT "Purification and characterization of a cadmium-induced
RT metallothionein from the shore crab Carcinus maenas (L.).";
RL Biochem. J. 297:609-614(1994).
CC -!- FUNCTION: Metallothioneins have a high content of cysteine
CC residues that bind various heavy metals. Class I MTS in marine
CC crustacea are involved in the sequestration of elevated levels of
CC heavy-metal ions.
CC -!- INDUCTION: By cadmium.
CC -!- SIMILARITY: Belongs to the metallothionein superfamily. Family 3.
DR PIR; S43367; S43367.
DR HSSP; P55949; IDMC.
DR InterPro; IPR003019; Metallothion_
DR InterPro; IPR002045; Metallothion_3.
DR Pfam; PF00131; Metallothio; 1.
DR PRINTS; PR00858; MTCRUSTACEAN.
KW Cadmium; Direct protein sequencing; Metal-binding;
KW Metal-thiolate cluster.
FT DOMAIN 1 29 Beta.
FT DOMAIN 30 58 Alpha.
FT METAL 5 5 Cluster B.
FT METAL 6 6 Cluster B.
FT METAL 10 10 Cluster B.
FT METAL 12 12 Cluster B.
FT METAL 17 17 Cluster B.
FT METAL 21 21 Cluster B.
FT METAL 23 23 Cluster B.
FT METAL 26 26 Cluster B.
FT METAL 28 28 Cluster B.
FT METAL 31 31 Cluster A.
FT METAL 34 34 Cluster A.
FT METAL 38 38 Cluster A.
FT METAL 40 40 Cluster A.
FT METAL 46 46 Cluster A.
FT METAL 50 50 Cluster A.
FT METAL 54 54 Cluster A.
FT METAL 56 56 Cluster A.
FT METAL 57 57 Cluster A.
FT VARIANT 1 1 Missing (in variant isolated in low
FT cadmium concentration).
SQ SEQUENCE 58 AA; 6133 MW;
0167CDA2E9C9731D CRC64;

Query Match 11.2%; Score 36.5; DB 1; Length 58;
Best Local Similarity 26.3%; Pred. No. 7.2e+03;
```

```

Matches 10; Conservative 2; Mismatches 11; Indels 15; Gaps 1;

QY 6 CAEKTFCVNGGECFMVKDLSNPSRYLCKPNEFTGDRG 43
Db 6 CIDKCECKEGG-----CKAGCKCTSCRC 28

RESULT 16
Q86M28
ID Q86M28 PRELIMINARY; PRT; 58 AA.
AC Q86M28;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Metallothionein Ib.
GN Name=MTIb;
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12058179;
RA Li B., Savva D.;
RT "Cloning of Metallothionein cDNAs and Its Gene in Shore Crab (Carcinus
maenas).";
RL Acta Biochim. Biophys. Sin. 32:640-644 (2000).
DR EMBL; AF196974; AAF07984.1; -.
DR HSSP; P55949; 1DME.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR InterPro; IPR003019; Metallothion_3.
DR InterPro; IPR002045; Metallothion_3.
DR InterPro; IPR001007; VWF C.
DR Pfam; PF00131; Metallothio; 1.
DR PRINTS; PR00858; MTCRUSTACEAN.
DR PROSITE; PS01208; VWF_C_1; UNKNOWN 1.
DR SEQUENCE 58 AA; 6186 MW; 0167CDA2F379731D CRC64;

Query Match 11.2%; Score 36.5; DB 2; Length 58;
Best Local Similarity 26.3%; Pred. No. 7.2e+03;
Matches 10; Conservative 2; Mismatches 11; Indels 15; Gaps 1;

QY 6 CAEKTFCVNGGECFMVKDLSNPSRYLCKPNEFTGDRG 43
Db 6 CIDKCECKEGG-----CKAGCKCTSCRC 28

RESULT 17
MT_ASTFL
ID MT_ASTFL STANDARD; PRT; 58 AA.
AC P55951;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Metallothionein (MT).
OS Astacus fluviatilis (Broad-fingered crayfish) (Astacus astacus).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Astacoidae; Astacidae; Astacus.
OX NCBI_TaxID=6715;
RN [1]
RP SEQUENCE, AND MASS SPECTROMETRY.
RC TISSUE=Midgut;
RX MEDLINE=97079279; PubMed=821011;
RA Pedersen S.N., Pedersen K.L., Hoejrup P., Depledge M.H., Knudsen J.;
RT "Primary structures of decapod crustacean metallothioneins with
special emphasis on freshwater and semi-terrestrial species.";
RL Biochem. J. 319:999-1003 (1996).
CC -!- FUNCTION: Metallothioneins have a high content of cysteine
residues that bind various heavy metals. Class I MTS in crustacea
are involved in the sequestration of elevated levels of heavy-
metal ions.

```

```

CC -!- INDUCTION: By cadmium.
CC -!- MASS SPECTROMETRY: MW=5910.8; METHOD=MALDI; RANGE=1-58;
CC NOTE=Ref.1.
DR -!- SIMILARITY: Belongs to the metallothionein superfamily. Family 3.
DR HSSP; P29499; 1J5L.
DR InterPro; IPR003019; Metallothion_3.
DR InterPro; IPR002045; Metallothion_3.
DR Pfam; PF00131; Metallothio; 1.
DR PRINTS; PR00858; MTCRUSTACEAN.
KW Cadmium; Direct protein sequencing; Metal-binding;
KW Metal-thiolate cluster.
FT DOMAIN 1 29 Beta.
FT DOMAIN 30 58 Alpha.
FT METAL 5 5 Cluster B.
FT METAL 6 6 Cluster B.
FT METAL 10 10 Cluster B.
FT METAL 12 12 Cluster B.
FT METAL 17 17 Cluster B.
FT METAL 21 21 Cluster B.
FT METAL 23 23 Cluster B.
FT METAL 26 26 Cluster B.
FT METAL 28 28 Cluster B.
FT METAL 31 31 Cluster A.
FT METAL 34 34 Cluster A.
FT METAL 38 38 Cluster A.
FT METAL 40 40 Cluster A.
FT METAL 46 46 Cluster A.
FT METAL 50 50 Cluster A.
FT METAL 54 54 Cluster A.
FT METAL 56 56 Cluster A.
FT METAL 57 57 Cluster A.
FT VARIANT 1 1 Missing (in variant isolated in low
cadmium concentration).
SQ SEQUENCE 58 AA; 5911 MW; 576365B3E5C7122 CRC64;

Query Match 11.0%; Score 36; DB 1; Length 58;
Best Local Similarity 33.3%; Pred. No. 8.3e+03;
Matches 10; Conservative 2; Mismatches 12; Indels 6; Gaps 2;

QY 5 KCAEKTFCVNGGECFMVKDLSNPSRYLCKC 34
Db 33 KCTSGCKPSKEC--AKTCSKP---CEC 56

RESULT 18
Q95U91
ID Q95U91 PRELIMINARY; PRT; 58 AA.
AC Q95U91;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Metallothionein.
OS Scylla serrata (Mud crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Portunoidae; Portunidae; Scylla.
OX NCBI_TaxID=6761;
RN [1]
RP SEQUENCE FROM N.A.
RL Li B., Fu X., Liu Q., Chen Y.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY057397; AAL23674.1; -.
DR HSSP; P29499; 1J5L.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR InterPro; IPR003019; Metallothion_3.
DR InterPro; IPR002045; Metallothion_3.
DR InterPro; IPR008197; WAP.
DR Pfam; PF00131; Metallothio; 1.
DR PRINTS; PR00003; 4DISULPHCORE.
DR PRINTS; PR00858; MTCRUSTACEAN.
SQ SEQUENCE 58 AA; 6239 MW; 8C66726F76E73341 CRC64;

Query Match 10.9%; Score 35.5; DB 2; Length 58;

```





RT New neurotoxin PRTx34C2 from venom of Brazilian Amazonian armed  
 RL spider phoreuxia reldyi";  
 CC Submitted (APR-2004) to Swiss-Prot.  
 CC -!- FUNCTION: No toxic effects on mice at dose levels of 5 ug per  
 CC mouse. May be toxic to insects.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.  
 CC -!- MASS SPECTROMETRY: MW=6330.3; METHOD=Electrospray; RANGE=1-58;  
 CC NOTE=Ref.1.  
 CC -!- SIMILARITY: Belongs to the spider toxin Tx3 family.  
 CC Calcium channel inhibitor; Direct protein sequencing;  
 KW Ionic channel inhibitor; Neurotoxin; Toxin.  
 SQ SEQUENCE 58 AA; 6341 MW; F4FF04A1FABCC49C CRC64;

Query Match 10.6%; Score 34.5; DB 1; Length 58;  
 Best Local Similarity 26.2%; Pred. No. 1.3e+04;  
 Matches 11; Conservative 3; Mismatches 11; Indels 17; Gaps 2;

Oy 12 CVNGGECFMVKDLSNPS:-----RYLC-----KCPN 36  
 Db 16 CCGNVCYCPDFWFGSQWPFCCSCAYGLKVCVCAHKQKCPN 57

RESULT 23  
 CTC7 HUMAN STANDARD; PRT; 58 AA.  
 AC Q9BQ2; 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Putative metallothionein C20orf127.  
 GN Name=C20orf127;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP MEDLINE=21638749; PubMed=11780052; DOI=10.1038/414865a;  
 RA Deloukas P., Matthews L.H., Ashurst J.L., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blake S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Lehaeslaio M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConachie L.J., McLeay K., McMurray A.A.,  
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaubin M., Wall M., Wallis J.M.,  
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 CC -!- FUNCTION: Metallothioneins have a high content of cysteine  
 CC residues that bind various heavy metals; these proteins are  
 CC transcriptionally regulated by both heavy metals and  
 CC glucocorticoids (By similarity).  
 CC -!- DOMAIN: Class I metallothioneins contain 2 metal-binding domains:  
 CC four divalent ions are chelated within cluster A of the alpha  
 CC domain and are coordinated via cysteinyl thiolate bridges to 11

CC cysteine ligands. Cluster B, the corresponding region within the  
 CC beta domain, can ligate three divalent ions to 9 cysteines.  
 CC -!- SIMILARITY: Belongs to the metallothionein superfamily. Family 1.  
 CC -!- CAUTION: Several lines of evidence, such as lack of introns,  
 CC truncation of the TATA box, lack of ESTs, suggest that it could be  
 CC a pseudogene.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AL121753; CAC34280.1; -;  
 DR HSSP; P02795; 1MHU.  
 DR Genew; HGNC:16120; C20orf127.  
 DR InterPro; IPR002400; GF\_cysknot.  
 DR InterPro; IPR003019; Metalthion\_1.  
 DR InterPro; IPR000006; Metalthion\_1.  
 DR InterPro; IPR001396; Metalthion\_4.  
 DR Pfam; PF00131; Metallothio; 1.  
 DR PRINTS; PR00438; GFCVSKNOT.  
 DR PRINTS; PR00873; MTECHINOIDEA.  
 DR PRINTS; PR00860; MIVTEBRATE.  
 DR PROSITE; PS00203; METALLOTHIONEIN VRT; 1.  
 KW Hypothetical protein; Metal-binding; Metal-thiolate cluster.  
 FT DOMAIN 1 29  
 FT DOMAIN 30 58  
 FT METAL 5 5  
 FT METAL 7 7  
 FT METAL 13 13  
 FT METAL 15 15  
 FT METAL 19 19  
 FT METAL 21 21  
 FT METAL 24 24  
 FT METAL 26 26  
 FT METAL 29 29  
 FT METAL 33 33  
 FT METAL 34 34  
 FT METAL 36 36  
 FT METAL 37 37  
 FT METAL 41 41  
 FT METAL 44 44  
 FT METAL 48 48  
 FT METAL 50 50  
 FT METAL 54 54  
 FT METAL 56 56  
 FT METAL 57 57  
 SQ SEQUENCE 58 AA; 5760 MW; D38G349E68A2DA3B CRC64;  
 Query Match 10.4%; Score 34; DB 1; Length 58;  
 Best Local Similarity 30.4%; Pred. No. 1.4e+04;  
 Matches 7; Conservative 0; Mismatches 8; Indels 8; Gaps 1;  
 Oy 12 CVNGGECFMVKDLSNPSRYLCCK 34  
 Db 7 CTTGSGCTCAGS-----CKC 21

RESULT 24  
 Q6K4G1 PRELIMINARY; PRT; 58 AA.  
 ID Q6K4G1  
 AC Q6K4G1  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Hypothetical protein OJ1451\_A02.25 (Hypothetical protein  
 DE P0499G10.6).  
 GN Name=OJ1451\_A02.25; Synonyms=P0499G10.6;  
 OS Oryza sativa (Japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;

```

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OX Ehrhartoideae; Oryzae; Oryza.
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Hattori M., Sasaki Y., Katayose Y.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005571; BAD23392.1; -.
DR EMBL; AP005587; BAD25995.1; -.
KW Hypothetical protein.
SQ SEQUENCE 58 AA; 5381 MW; F1265905C0B3E864 CRC64;

Query Match 10.4%; Score 34; DB 2; Length 58;
Best Local Similarity 25.0%; Pred. No. 1.4e+04;
Matches 9; Conservative 9; Mismatches 16; Indels 2; Gaps 1;

Qy 22 KDLNPSRYLCKPNEFTGDR--QNYVMASFYKAE 55
Db 22 EDLQAFDVMGALPDVFSGGHTCEDFVRAIIHGAK 57

RESULT 25
Q8GU49 PRELIMINARY; PRT; 58 AA.
ID Q8GU49
AC Q8GU49;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative metallothionein (Fragment).
GN Name=mt2a;
OS Posidonia oceanica.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Posidoniaceae; Posidonia.
OX NCBI_TaxID=55489;
RN [1]
RP SEQUENCE FROM N.A.
RA Maestri P., Giordani T., Natali L., Cavallini A.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ311342; CAC85330.1; -.
DR GO; GO:0046872; F:metal ion binding; IEA.
FT NON_TER 1
FT NON_TER 58
SQ SEQUENCE 58 AA; 5863 MW; E2B5D97DDB20FCBC CRC64;

Query Match 10.2%; Score 33.5; DB 2; Length 58;
Best Local Similarity 45.0%; Pred. No. 1.6e+04;
Matches 9; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

Qy 6 CAEKTFCVNG-GECFPMVKDL 24
Db 3 CGSGCNCNGCGCGCMYFDL 22

RESULT 26
Q8H494 PRELIMINARY; PRT; 58 AA.
ID Q8H494
AC Q8H494;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein P0427D10.141 (Hypothetical protein
DE OSJNBa0084D17.3).
CN Name=P0427D10.141; Synonyms=OSJNBa0084D17.3;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]

RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
RT clone:OSJNBa0084D17.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004272; BAC20064.1; -.
DR EMBL; AP006459; BAC32036.1; -.
DR Gramene; Q8H494; -.
KW Hypothetical protein.
SQ SEQUENCE 58 AA; 6429 MW; ODA2EDA130F9ED2E CRC64;

Query Match 10.2%; Score 33.5; DB 2; Length 58;
Best Local Similarity 25.0%; Pred. No. 1.6e+04;
Matches 10; Conservative 9; Mismatches 18; Indels 3; Gaps 2;

Qy 15 GGECFMVKDLNPSRYLC-KCPNEFTGDRG--QNYVMASF 51
Db 5 GYRCQVIESVGNAGTILAFTCATVVLCEQCQYQHIVAIF 44

RESULT 27
Q9QA93 PRELIMINARY; PRT; 58 AA.
ID Q9QA93
AC Q9QA93;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Helicase (Fragment).
GN Name=NS3;
OS GB virus C/Hepatitis G virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC GBV-C/HGV group.
OX NCBI_TaxID=54290;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=69251;
RC MEDLINE=97306727; PubMed=9163994;
RX Pinho J.R., Capacci M.L., da Silva L.C., Carrilho F.J., Santos C.A.,
RA Pugliese V., Guz B., Levi J.E., Ballarati C.A., Bernardini A.P.;
RA "Hepatitis G virus/GB virus C in Brazil. Preliminary report.";
RL Rev. Inst. Med. Trop. Sao Paulo 38:243-246(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=69251;
RC MEDLINE=99221812; PubMed=10203545;
RX Pinho J.R.R., Zanotto P.M., Ferreira J.L.P., Sumita L.M.,
RA Carrilho F.J., da Silva L.C., Capacci M.L., Silva A.O., Guz B.,
RA Goncalves Jr F.L., Goncalves N.S.L., Buck G.A., Meyers G.A.,
RA Bernardini A.P.;
RA "High prevalence of GB virus C in Brazil and molecular evidence for
RT intrafamilial transmission.";
RL J. Clin. Microbiol. 37:1634-1637(1999).
DR EMBL; AF124773; AAF20306.1; -.
DR HSSP; P28663; 8OHM.
DR MEROPS; S29.002; -.
DR GO; GO:0004386; F:helicase activity; IEA.
KW Helicase.
FT NON_TER 1
FT NON_TER 58
SQ SEQUENCE 58 AA; 6493 MW; 19FB427C9B5226F6 CRC64;

Query Match 10.2%; Score 33.5; DB 2; Length 58;
Best Local Similarity 31.4%; Pred. No. 1.6e+04;
Matches 11; Conservative 1; Mismatches 12; Indels 11; Gaps 1;

Qy 2 HLVKCAEKTFC-----VNGGCFPMVKDL 25
Db 17 HLVFCHSKAECERLAGQFSARGVNAAXAYRGKSS 51

RESULT 28
Q6A1C0 PRELIMINARY; PRT; 58 AA.
ID Q6A1C0
```

AC Q6ALC0;  
 DT 25-OCT-2004 (TRENBLrel. 28, Created)  
 DT 25-OCT-2004 (TRENBLrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TRENBLrel. 28, Last annotation update)  
 DE Putative albumin 1 (Fragment).  
 GN Name=pal;  
 OS Vigna subterranea.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.  
 OX NCBI\_TaxID=115715;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Higgins T.J., Chandler P.M., Randall P.J., Spencer D., Beach L.R.,  
 RA Blagrove R.J., Kortt A.A., Inglis A.S.;  
 RT "Gene structure, protein structure, and regulation of the synthesis of  
 RT a sulfur-rich protein in pea seeds.";  
 RL J. Biol. Chem. 261:11124-11130(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Louis S., Delobel B., Gressent F., Rahioui I., Dupont G., Diol O.,  
 RA Chessel D., Rahbe Y.;  
 RT "Broad screening of the legume family for variability of insecticidal  
 RT activities and occurrence of seed albumin 1b toxins.";  
 RL Submitted (JUL-2004) to the EMBL/genbank/DBJ databases.  
 DR EMBL; AJ784958; CAH05258.1; -  
 FT NON\_TER 1 1  
 FT CHAIN <1 22 putative albumin 1b.  
 FT CHAIN 31 >58 putative albumin 1a.  
 FT NON\_TER 58 58  
 FT NON\_TER 58 58  
 SQ SEQUENCE 58 AA; 6345 MW; 73B7146B8C60F169 CRC64;  
 Query Match 10.1%; Score 33; DB 2; Length 58;  
 Best Local Similarity 26.8%; Pred. No. 1.9e+04;  
 Matches 11; Conservative 6; Mismatches 14; Indels 10; Gaps 2;  
 QY 16 GECFMVKDLSNPSRY-----LC-----KCPNEFTGDRQNY 46  
 Db 15 GFCIYPTGLSSVKMIDHPNLQSHDECMKKSGNFCARY 55  
 RESULT 29  
 Q8XPP2  
 ID Q8XPP2 PRELIMINARY; PRT; 58 AA.  
 AC Q8XPP2;  
 DT 01-MAR-2002 (TRENBLrel. 20, Created)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Hypothetical protein Rsp1596.  
 GN Name=RS02153; OrderedLocusNames=RSpl596;  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Plasmid megaplasmid.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681879; PubMed=11823852; DOI=10.1038/415497a;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
 RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 RA Siguiet P., Thebault P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";  
 RL Nature 415:497-502(2002).  
 DR EMBL; AL646085; CAD18747.1; -  
 DR Pfam; PF07519; Tannase; 1.  
 KW Complete proteome; Hypothetical protein; Plasmid.  
 SQ SEQUENCE 58 AA; 6100 MW; 4A5BEE1BECSD7717 CRC64;  
 Query Match 10.1%; Score 33; DB 2; Length 58;

Best Local Similarity 22.4%; Pred. No. 1.9e+04;  
 Matches 11; Conservative 7; Mismatches 9; Indels 22; Gaps 2;  
 QY 6 CAEKTFCVNGGECFMVKDLSNPSRYLKCP-----NEFTGDRC 43  
 Db 20 CADRNVAHG-----RTRPLCRYPKVARRTGAGDLNDASGLRC 57  
 RESULT 30  
 Q97JPA  
 ID Q97JPA PRELIMINARY; PRT; 58 AA.  
 AC Q97JPA;  
 DT 01-OCT-2001 (TRENBLrel. 18, Created)  
 DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Hypothetical protein CAC1229.  
 GN OrderedLocusNames=CAC1229;  
 OS Clostridium acetobutylicum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 824 / DSM 792 / VGM B-1787;  
 RX MEDLINE=21359325; PubMed=11466286;  
 RX DOI=10.1128/JB.183.16.4823-4838.2001;  
 RA Neolling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L.A., Soucaille P.,  
 RA Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.;  
 RT "Genome sequence and comparative analysis of the solvent-producing  
 RT bacterium Clostridium acetobutylicum.";  
 RL J. Bacteriol. 183:4823-4838(2001).  
 DR EMBL; AE007636; AAK79201.1; -  
 DR PIR; F97051; F97051.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 58 AA; 6885 MW; 0D2C3F818482AF99 CRC64;  
 Query Match 10.1%; Score 33; DB 2; Length 58;  
 Best Local Similarity 83.3%; Pred. No. 1.9e+04;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 41 DRQNY 46  
 Db 25 DRCQNW 30  
 Search completed: April 14, 2005, 08:39:37  
 Job time : 185 secs

This page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2005, 08:28:21 ; Search time 42 Seconds  
(without alignments)  
132.871 Million cell updates/sec

Title: US-09-980-672-2  
Perfect score: 327  
Sequence: 1 SHLVKCAEKTFCVNGGSCFM.....TGDRCONVMAFYKAEELY 58

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 356

Minimum DB seq length: 58  
Maximum DB seq length: 58

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : p1r1: 1  
p1r2: 2  
p1r3: 3  
p1r4: 4

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43.5	13.3	58	2 A37039	metallothionein 1
2	39	11.9	58	1 SKMD1S	metallothionein 1
3	38.5	11.8	58	2 S59072	metallothionein is
4	36.5	11.2	58	2 S43367	metallothionein -
5	33	10.1	58	2 F97051	hypothetical prote
6	32.5	9.9	58	2 S71422	argininosuccinate
7	32.5	9.9	58	2 F75180	hypothetical prote
8	32	9.8	58	2 B97333	ferredoxin (import
9	32	9.8	58	2 A96017	hypothetical prote
10	31	9.5	58	2 D49038	Ig lambda chain V
11	31	9.5	58	2 G36953	hypothetical prote
12	30	9.2	58	2 G37810	hypothetical prote
13	30	9.2	58	2 H64377	ferredoxin 2[4Fe-4
14	29.5	9.0	58	2 F71334	conserved hypothet
15	28	8.6	58	2 JG6527	antibiotic mutaci
16	28	8.6	58	2 G31033	hypothetical prote
17	28	8.6	58	2 D2970	hypothetical prote
18	28	8.6	58	2 H85877	hypothetical prote
19	27.5	8.4	58	2 C69365	ferredoxin (fdx-5)
20	27	8.3	58	2 D95214	hypothetical prote
21	27	8.3	58	2 D48652	transfer protein s
22	27	8.3	58	2 F64318	hypothetical prote
23	27	8.3	58	2 B59060	attractin - sooty
24	27	8.3	58	2 A59060	attractin - Califo
25	27	8.3	58	2 C98078	hypothetical prote
26	27	8.3	58	2 A59447	attractin - Aplysi
27	27	8.3	58	2 B42960	ferredoxin 2[4Fe-4
28	26.5	8.1	58	2 S10063	isoaprotinin G2 -
29	26.5	8.1	58	2 I37558	RNA polymerase II

30	26.5	8.1	58	2 C83918	hypothetical prote
31	26.5	8.1	58	2 D89984	hypothetical prote
32	26	8.0	58	2 I77464	luteinizing hormon
33	26	8.0	58	2 H86595	ct702 hypothetical
34	26	8.0	58	2 B72028	hypothetical prote
35	26	8.0	58	2 C82527	hypothetical prote
36	25.5	7.8	58	2 D82057	ribosomal protein
37	25.5	7.8	58	2 B95235	preprotein translo
38	25.5	7.8	58	2 C75890	hypothetical prote
39	25.5	7.8	58	2 C98099	preprotein translo
40	25	7.6	58	1 N1EP2J	short neurotoxin 2
41	25	7.6	58	2 F84303	50S ribosomal prot
42	25	7.6	58	2 S77421	glycoprotein 64 -
43	25	7.6	58	2 G91284	hypothetical prote
44	25	7.6	58	2 S73144	hypothetical prote
45	25	7.6	58	2 E97047	hypothetical prote
46	24.5	7.5	58	2 G81707	hypothetical prote
47	24	7.3	58	1 LBRF1S	light-harvesting p
48	24	7.3	58	1 LBRPAC	Ig heavy chain V r
49	24	7.3	58	2 S03064	zinc finger protei
50	24	7.3	58	2 A43284	glycophorin M11 -
51	24	7.3	58	2 I52571	hypothetical prote
52	24	7.3	58	2 AC1517	hypothetical prote
53	24	7.3	58	2 AF1158	methionine adenosy
54	23.5	7.2	58	2 S77764	SSU ribosomal prot
55	23.5	7.2	58	2 E69416	histone H3 - mouse
56	23	7.0	58	2 S45111	protein YDR034C-a
57	23	7.0	58	2 D56613	virion morphogenes
58	23	7.0	58	2 S78711	prophage pi2 prote
59	23	7.0	58	2 H86755	hypothetical prote
60	23	7.0	58	2 E81202	hypothetical prote
61	23	7.0	58	2 C69382	hypothetical prote
62	23	7.0	58	2 D82759	hypothetical prote
63	23	7.0	58	2 A97130	hypothetical prote
64	23	7.0	58	2 D97842	hypothetical prote
65	23	7.0	58	2 G97818	hypothetical prote
66	23	7.0	58	2 H95384	protein (imported
67	22.5	6.9	58	1 Z5BP83	gene 58 protein -
68	22.5	6.9	58	2 A82541	hypothetical prote
69	22	6.7	58	2 PX0017	glutathione-disulf
70	22	6.7	58	2 S35573	zona pellucida gly
71	22	6.7	58	2 S16486	rRNA N-glycosidase
72	22	6.7	58	2 T16505	hypothetical prote
73	22	6.7	58	2 G44840	GDFLRamide neuro
74	22	6.7	58	2 T07265	hypothetical prote
75	22	6.7	58	2 H84534	hypothetical prote
76	22	6.7	58	2 H83198	hypothetical prote
77	22	6.7	58	2 T10833	Y4xF protein - Rhi
78	22	6.7	58	2 C81896	hypothetical prote
79	22	6.7	58	2 T29075	hypothetical prote
80	22	6.7	58	2 F69092	conserved hypothet
81	22	6.7	58	2 C82803	hypothetical prote
82	22	6.7	58	2 C97788	hypothetical prote
83	22	6.7	58	2 AD2317	hypothetical prote
84	21.5	6.6	58	2 B81983	hypothetical prote
85	21.5	6.6	58	2 E57256	hypothetical prote
86	21.5	6.6	58	2 S58628	hypothetical prote
87	21	6.4	58	1 WMTM68	6.8K protein - tob
88	21	6.4	58	2 A71604	ribosomal protein
89	21	6.4	58	2 S03810	psal protein - Rhi
90	21	6.4	58	2 B44358	cysteine-rich prot
91	21	6.4	58	2 E64021	hypothetical prote
92	21	6.4	58	2 D86535	hypothetical prote
93	21	6.4	58	2 B82454	hypothetical prote
94	21	6.4	58	2 B69871	hypothetical prote
95	21	6.4	58	2 H69930	hypothetical prote
96	21	6.4	58	2 S73287	hypothetical prote
97	21	6.4	58	2 A72088	hypothetical prote
98	21	6.4	58	2 A82628	hypothetical prote
99	21	6.4	58	2 E97793	hypothetical prote
100	21	6.4	58	2 C46435	Na+/K+-exchanging

## ALIGNMENTS

## RESULT 1

metallothionein 1 - American lobster  
C;Species: Homarus americanus (American lobster)  
C;Date: 31-Jan-1992 #sequence\_revision 31-Jan-1992 #text\_change 09-Jul-2004  
C;Accession: A37039  
R;Brouwer, M.; Winge, D.R.; Gray, W.R.  
J. Inorg. Biochem. 35, 289-303, 1989  
A;Title: Structural and functional diversity of copper-metallothioneins from the Americas  
A;Reference number: A37039; MUID:89215793; PMID:2709004  
A;Accession: A37039  
A;Status: preliminary  
A;Molecule type: protein  
A;Residues: 1-58 <PRO>  
A;Cross-references: UNIPROT:P29499  
C;Superfamily: metallothionein

Query Match 13.3%; Score 43.5; DB 2; Length 58;  
Best Local Similarity 25.0%; Pred. No. 3.1e+02;  
Matches 12; Conservative 6; Mismatches 13; Indels 17; Gaps 3;

QY 6 CAETFCVNGG-----ECFVVKDLSNPRLKCPNEFTGDRG 43

DB 5 CKDKCEAEGGCKTGCKTSCRCAPCEKCTSG----CKPSK---DEC 45

## RESULT 2

SMKDI5  
metallothionein 1 - mud crab  
C;Species: Scylla serrata (mud crab)  
C;Date: 29-Jul-1981 #sequence\_revision 29-Jul-1981 #text\_change 09-Jul-2004  
C;Accession: A03283  
R;Lerch, K.; Ammer, D.; Olafson, R.W.  
J. Biol. Chem. 257, 2420-2426, 1982  
A;Title: Crab metallothionein. Primary structures of metallothioneins 1 and 2.  
A;Reference number: A92363; MUID:82142340; PMID:7061431  
A;Accession: A03283  
A;Molecule type: protein  
A;Residues: 1-58 <LER>  
A;Cross-references: UNIPROT:P02805  
A;Note: The five Cys-X-Cys sequences are believed to be the principal metal-binding sites  
C;Superfamily: metallothionein  
C;Keywords: metal binding

Query Match 11.9%; Score 39; DB 1; Length 58;  
Best Local Similarity 26.2%; Pred. No. 1e+03;  
Matches 11; Conservative 4; Mismatches 13; Indels 14; Gaps 2;

QY 6 CAETFCVNGG-----ECFVVKDLSNPRLKCPNE 37

DB 5 CNDKVCVKEGCKEGCQCTSCRCSPCEKCSG----CKANK 42

## RESULT 3

S59072  
metallothionein isoform Ia - blue crab  
C;Species: Callinectes sapidus (blue crab)  
C;Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 09-Jul-2004  
C;Accession: S59072  
R;Brouwer, M.; Enghild, J.; Hoexum-Brouwer, T.; Thøgersen, I.; Truncali, A.  
Biochem. J. 311, 617-622, 1995  
A;Title: Primary structure and tissue-specific expression of blue crab (Callinectes sapidus) metallothionein  
A;Reference number: S59072; MUID:96033062; PMID:7487904  
A;Accession: S59072  
A;Molecule type: protein  
A;Residues: 1-58 <PRO>  
A;Cross-references: UNIPROT:P55949  
C;Superfamily: metallothionein  
C;Keywords: metal binding

## Query Match

Best Local Similarity 11.8%; Score 38.5; DB 2; Length 58;  
Matches 10; Conservative 2; Mismatches 11; Indels 15; Gaps 1;

QY 6 CAETFCVNGGECFVVKDLSNPRLKCPNEFTGDRG 43

DB 5 CNDKVCVKEG-----CKAGCQCTSCRC 27

## RESULT 4

S43367  
metallothionein - green crab  
C;Species: Carcinus maenas (green crab, common shore crab)  
C;Date: 19-Mar-1997 #sequence\_revision 01-Feb-1999 #text\_change 09-Jul-2004  
C;Accession: S43367  
R;Pedersen, K.L.; Pedersen, S.N.; Hojrup, P.; Andersen, J.S.; Roepstorff, P.; Knudsen, J.  
Biochem. J. 297, 609-614, 1994  
A;Title: Purification and characterization of a cadmium-induced metallothionein from the green crab, *Carcinus maenas*  
A;Reference number: S43367; MUID:94153337; PMID:8110201  
A;Accession: S43367  
A;Molecule type: protein  
A;Residues: 1-58 <PEB>  
A;Cross-references: UNIPROT:P55948  
A;Note: The sequence from Fig. 4 is inconsistent with that from Fig. 6 in having an additional Cys at position 58  
C;Superfamily: metallothionein  
C;Keywords: metal binding; chelation; metal-thiolate cluster  
F;6.12.17.21.28/Binding site: transition metal ions (Cys) #status predicted  
F;3.38.46.47.54.56.57/Binding site: transition metal ions (Cys) #status predicted

Query Match 11.2%; Score 36.5; DB 2; Length 58;  
Best Local Similarity 26.3%; Pred. No. 2e+03;  
Matches 10; Conservative 2; Mismatches 11; Indels 15; Gaps 1;

QY 6 CAETFCVNGGECFVVKDLSNPRLKCPNEFTGDRG 43

DB 6 CIDKCECKEG-----CKAGCQCTSCRC 28

## RESULT 5

F97051  
hypothetical protein CAC1229 [imported] - Clostridium acetobutylicum  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C;Accession: F97051  
R;Nolling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Baly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4836, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum* ATCC8224  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: F97051  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-58 <KUR>  
A;Cross-references: UNIPROT:Q97JP4; GB:AE001437; PID:AAK79201.1; PID:gl5024154; GSPDB:GB:AE001437  
A;Experimental source: Clostridium acetobutylicum ATCC8224  
C;Genetics:  
A;Gene: CAC1229

Query Match 10.1%; Score 33; DB 2; Length 58;  
Best Local Similarity 83.3%; Pred. No. 5e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 41 DRCONY 46

DB 25 DRCONW 30

## RESULT 6

S71422  
argininosuccinate lyase - duck (fragments)  
N;Alternate names: Delta 2 crystallin  
C;Species: Anas platyrhynchos (domestic duck)  
C;Date: 19-Mar-1998 #sequence\_revision 17-Apr-1998 #text\_change 09-Jul-2004

Best local similarity	Seq. no.	Indels	Gaps
9; Conservative	32; 18	4; Mismatches	12; Gaps
9; Conservative	32; 18	4; Mismatches	12; Gaps

hypothetical protein (acol 3' region) - Pelobacter carbinolicus  
 C:Species: Pelobacter carbinolicus  
 C:Date: 07-Jul-1995 #sequence\_revision 07-Jul-1995 #text\_change 12-Jul-2004  
 C:Accession: G36953  
 R:Oppermann, F.B.; Steinbuechel, A.  
 J. Bacteriol. 176, 469-485, 1994  
 A:Title: Identification and molecular characterization of the aco genes encoding the Pel  
 A:Reference number: A36953; MUID:94117383; PMID:8110297  
 A:Accession: G36953  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-58 <OPP>  
 A:Cross-references: UNIPROT:Q51596; GB:U01100; NID:g434021; PIDN:AAA18920.1; PID:g434028

Query Match 9.5%; Score 31; DB 2; Length 58;  
 Best Local Similarity 45.5%; Pred. No. 8.4e+03; Mismatches 2; Indels 0; Gaps 0;  
 Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 8 EKTFCVNGGEC 18  
 :|||  
 Db 7 DKSRCTQCCEC 17  
 :|||

RESULT 12  
 G97810  
 conserved hypothetical protein RC0887 [imported] - Rickettsia conorii (strain Malish 7)  
 C:Species: Rickettsia conorii  
 C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 09-Jul-2004  
 C:Accession: G97810  
 R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R  
 Science 293, 2093-2098, 2001  
 A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
 A:Reference number: A57700; MUID:21442074; PMID:11557893  
 A:Accession: G97810  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-58 <KUR>  
 A:Cross-references: UNIPROT:Q92H84; GB:AE006914; PIDN:AAL03425.1; PID:gi5619993; GSPDB:G  
 C:Genetics:  
 A:Gene: RC0887

Query Match 9.2%; Score 30; DB 2; Length 58;  
 Best Local Similarity 35.3%; Pred. No. 1.1e+04; Mismatches 1; Indels 0; Gaps 0;  
 Matches 6; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

Qy 31 LCKCPNEFTGRCQNYV 47  
 :|||  
 Db 38 ICKCLGAIATVCCNYV 54  
 :|||

RESULT 13  
 H64377  
 ferredoxin 2[4Fe-4S] - Methanococcus jannaschii  
 C:Species: Methanococcus jannaschii  
 C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 12-Jul-2004  
 C:Accession: H64377  
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
 Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;  
 rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
 Science 273, 1058-1073, 1996  
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C  
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
 A:Reference number: A64300; MUID:96337999; PMID:8688087  
 A:Accession: H64377  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-58 <BUL>  
 A:Cross-references: UNIPROT:Q58041; GB:U67510; GB:L77117; NID:gi591325; PIDN:AAB98619.1;  
 C:Genetics:  
 A:Map position: FOR553382-553558  
 C:Keywords: 4Fe-4S; electron transfer; iron-sulfur protein; metalloprotein  
 F:1-55/Domain: ferredoxin 2[4Fe-4S] homology <FER3>  
 F:9,12,15,47/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

F:19,37,40,43/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 9.2%; Score 30; DB 2; Length 58;  
 Best Local Similarity 46.7%; Pred. No. 1.1e+04; Mismatches 1; Indels 2; Gaps 1;  
 Matches 7; Conservative 1; Mismatches 5; Indels 2; Gaps 1;

Qy 4 VKCAERTFCVNGGEC 18  
 :|||  
 Db 3 IKILEK--CVGCGNC 15  
 :|||

RESULT 14  
 F71334  
 conserved hypothetical protein TP0360 - syphilis spirochete  
 C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
 C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 09-Jul-2004  
 C:Accession: F71334  
 R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin  
 rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo  
 rthey, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
 Science 281, 375-388, 1998  
 A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
 A:Reference number: A71250; MUID:98332770; PMID:9665876  
 A:Accession: F71334  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-58 <COL>  
 A:Cross-references: UNIPROT:O83379; GB:AE001215; GB:AE000520; NID:g3322631; PIDN:AAC6534;  
 A:Experimental source: strain Nichols  
 C:Genetics:  
 A:Gene: TP0360

Query Match 9.0%; Score 29.5; DB 2; Length 58;  
 Best Local Similarity 22.2%; Pred. No. 1.2e+04; Mismatches 7; Indels 1; Gaps 1;  
 Matches 8; Conservative 7; Mismatches 20; Indels 1; Gaps 1;

Qy 1 SHLVKCAETFCVNGGECFMVKDLNFS-RYLCKCP 35  
 :|||  
 Db 10 SRTAAATQRFNCPCGGEVLRISIVDNGKVKNIACEP 45  
 :|||

RESULT 15  
 JC6527  
 Lantibiotic mutacin II protein MutM - Streptococcus mutans  
 C:Species: Streptococcus mutans  
 C:Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 17-Mar-1999  
 C:Accession: JC6527  
 R:Woodruff, W.A.; Novak, J.; Caufield, P.W.  
 Gene 206, 37-43, 1998  
 A:Title: Sequence analysis of mutA and mutM genes involved in the biosynthesis of the lan  
 A:Reference number: JC6526; MUID:98121310; PMID:9461412  
 A:Accession: JC6527  
 A:Molecule type: mRNA  
 A:Residues: 1-58 <WOO>  
 A:Cross-references: GB:U40620  
 C:Comment: This protein is the modification enzyme catalyzing dehydration and the formati  
 C:Genetics:  
 A:Gene: mutM

Query Match 8.6%; Score 28; DB 2; Length 58;  
 Best Local Similarity 25.7%; Pred. No. 1.8e+04; Mismatches 4; Conservative 6; Mismatches 5; Indels 0; Gaps 0;  
 Matches 4; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 44 QNYWASFYKAEELY 58  
 :|||  
 Db 29 ENFIDISIFKLIKY 43  
 :|||

RESULT 16  
 H91033  
 Hypothetical protein ECs3239 [imported] - Escherichia coli (strain O157:H7, substrain RIN  
 C:Species: Escherichia coli  
 C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004



C;Accession: G91033  
R;Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
Sasawara, N.; Yasunaga, T.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic reference number: A99629; MUID:21156231; PMID:11258796  
A;Accession: G91033  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-58 <HAY>  
A;Cross-references: UNIPROT:Q8XCL5; GB:BA000007; PIDN:BA036662.1; PID:g13362709; GSPDB:G  
A;Experimental source: strain O157:H7, substrain RMD 0509952  
C;Genetics:  
A;Gene: ECs3239

Query Match 8.6%; Score 28; DB 2; Length 58;  
Best Local Similarity 43.8%; Pred. No. 1.8e+04;  
Matches 7; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 20 MVKDLNPSRYLCKCP 35  
Db 1 MEKDLKELREYLLSP 16

RESULT 17  
D82970  
hypothetical protein PA5408 [imported] - *Pseudomonas aeruginosa* (strain PA01)  
C;Species: *Pseudomonas aeruginosa*  
C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004  
A;Accession: D82970  
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho  
A;Reference number: A82950; MUID:20437337; PMID:10984043  
A;Accession: D82970  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-58 <STO>  
A;Cross-references: UNIPROT:Q9HTF6; GB:AE004953; GB:AE004091; NID:g9951727; PIDN:AAG0879  
A;Experimental source: strain PA01  
C;Genetics:  
A;Gene: PA5408

Query Match 8.6%; Score 28; DB 2; Length 58;  
Best Local Similarity 37.0%; Pred. No. 1.8e+04;  
Matches 10; Conservative 5; Mismatches 6; Indels 6; Gaps 2;

QY 5 KCAETFCV-----NGGECFMVKDL 25  
Db 4 KPAETRCWLRQDDNGNAFVRRDLT 30

RESULT 18  
H85877  
hypothetical protein Z3621 [imported] - *Escherichia coli* (strain O157:H7, substrain EDL9  
C;Species: *Escherichia coli*  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
A;Accession: H85877  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: H85877  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-58 <STO>  
A;Cross-references: UNIPROT:Q8XCL5; GB:AE005174; NID:g12516726; PIDN:AGC57484.1; GSPDB:G  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: Z3621

Query Match 8.6%; Score 28; DB 2; Length 58;  
Best Local Similarity 43.8%; Pred. No. 1.8e+04;  
Matches 7; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 20 MVKDLNPSRYLCKCP 35  
Db 1 MEKDLKELREYLLSP 16

RESULT 19  
C69365  
ferredoxin (fdx-5) homolog - *Archaeoglobus fulgidus*  
C;Species: *Archaeoglobus fulgidus*  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 12-Jul-2004  
A;Accession: C69365  
R;Klonk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson,  
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;  
Glocke, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.N.  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
A;Reference number: A69250; MUID:98049343; PMID:9389475  
A;Accession: C69365  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-58 <KLE>  
A;Cross-references: UNIPROT:Q29339; GB:AE001040; GB:AE000782; NID:g2689363; PIDN:AAB9032;  
F;3-57/Domain: ferredoxin 2[4Fe-4S] homology <PER>

Query Match 8.4%; Score 27.5; DB 2; Length 58;  
Best Local Similarity 20.0%; Pred. No. 2.1e+04;  
Matches 8; Conservative 6; Mismatches 11; Indels 15; Gaps 2;

QY 5 KCAETFCVNGGECFMVKDLNPSRYLC-----KCPN 36  
Db 19 ECKGNAIHLGSKAYVDED-----ACMEGACVRACP 51

RESULT 20  
D95214  
hypothetical protein SP1836 [imported] - *Streptococcus pneumoniae* (strain TIGR4)  
C;Species: *Streptococcus pneumoniae*  
C;Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
A;Accession: D95214  
R;Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, E.  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A;Title: Complete Genome Sequence of a virulent isolate of *Streptococcus pneumoniae*.  
A;Reference number: A95000; MUID:21357209; PMID:11463916  
A;Accession: D95214  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-58 <KUR>  
A;Cross-references: UNIPROT:Q97P08; GB:AE005672; PIDN:AAK75909.1; PID:g14973337; GSPDB:G  
A;Experimental source: strain TIGR4  
C;Genetics:  
A;Gene: SP1836

Query Match 8.3%; Score 27; DB 2; Length 58;  
Best Local Similarity 20.0%; Pred. No. 2.3e+04;  
Matches 3; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 42 RCQNYVMASFYKAE 56  
Db 8 KCYTKLLTSYFKPRD 22

RESULT 21  
D48652



submitted to the Protein Sequence Database, September 2002

A:Description: .  
A:Reference number: A59447  
A:Accession: A59447  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-58 <NAG>  
A:Cross-references: UNIPROT:Q7M459  
A:Experimental source: albumen gland  
A>Note: Water-borne peptide pheromonal attractant  
C:Keywords: glycoprotein; pheromone  
F:1-58/Product: attractin #status experimental <MAT>  
F:4-41,13-33,20-26/Disulfide bonds: #status predicted  
F:8/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 8.3%; Score 27; DB 2; Length 58;  
Best Local Similarity 80.0%; Pred. No. 2.3e+04;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 41 DRQCN 45  
DB 39 ERCQN 43

RESULT 27  
B42960  
ferredoxin 2[4Fe-4S] homolog - Methanosarcina thermophila  
C:Species: Methanosarcina thermophila  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 12-Jul-2004  
C:Accession: B42960  
R:Clements, A.P.; Perry, J.G.  
J. Bacteriol. 174, 5244-5250, 1992  
A:Title: Cloning, nucleotide sequence, and transcriptional analyses of the gene encoding  
A:Reference number: A42960; MUID:92355496; PMID:1379583  
A:Contents: TM-1  
A:Accession: B42960  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-58 <CLE>  
A:Cross-references: UNIPROT:Q01700; GB:M83188; NID:g149823; PIDN:AAA73172.1; PID:g149824  
A>Note: sequence extracted from NCBI backbone (NCBIN:110325, NCBI:P110326)  
C:Keywords: 4Fe-4S; iron-sulfur protein; metalloprotein  
F:3-58/Domain: ferredoxin 2[4Fe-4S] homolog <FER>  
F:10,13,16,50/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
F:20,40,43,46/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 8.3%; Score 27; DB 2; Length 58;  
Best Local Similarity 21.2%; Pred. No. 2.3e+04;  
Matches 7; Conservative 7; Mismatches 7; Indels 12; Gaps 2;

QY 4 VKCAEKTFCVNGGECFMVKDLSNPSRYLCKCPN 36  
DB 32 IATVDESECLDCGSC---ED-----ACPN 52

RESULT 28  
S10063  
isoprotinin G2 - bovine hybrid  
C:Species: Bos primigenius indicus x Bos primigenius taurus (cattle)  
C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 09-Jul-2004  
C:Accession: S10063  
R:Siekman, J.; Wenzel, H.R.; Schroeder, W.; Tschesche, H.  
Biol. Chem. Hoppe-Seyler 369, 157-163, 1988  
A:Title: Characterization and sequence determination of six aprotinin homologues from bo  
A:Reference number: S00371; MUID:88221840; PMID:2453200  
A:Accession: S10063  
A:Molecule type: protein  
A:Residues: 1-58 <SIE>  
A:Cross-references: UNIPROT:Q7M312  
C:Superfamily: basic proteinase inhibitor; animal Kunitz-type proteinase inhibitor homol  
C:Keywords: serine proteinase inhibitor  
F:5-55/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>  
F:15/Inhibitory site: Lys (trypsin) #status predicted

Query Match 8.1%; Score 26.5; DB 2; Length 58;  
Best Local Similarity 27.5%; Pred. No. 2.7e+04;  
Matches 14; Conservative 8; Mismatches 20; Indels 9; Gaps 4;

QY 11 FCVN---GGEQ--PMVKDLSNPSRYLCKCPNEFTGDRCONYVMASFYKAE 56  
DB 4 FCLEPPYTGPCAKMIRYFYNAKGLCO-PFVYGCRKR---NNFKSAED 50

## RESULT 29

I37558

RNA polymerase II - human

C:Species: Homo sapiens (man)

C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 09-Jul-2004

C:Accession: I37558; S53014

R:Shpakovski, G.V.; Acker, J.; Wintzerith, M.; Lacroix, J.F.; Thuriaux, P.; Vigneron, M.  
Mol. Cell. Biol. 15, 4702-4710, 1995A:Title: Four subunits that are shared by the three classes of RNA polymerase are function  
A:Reference number: I37558; MUID:95379812; PMID:7651387

A:Accession: I37558

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-58 &lt;RES&gt;

A:Cross-references: UNIPROT:P53803; EMBL:Z47727; NID:g717186; PIDN:CAA87656.1; PID:g71718  
R:Vigneron, M.

submitted to the EMBL Data Library, January 1995

A:Description: Two human polypeptides are functionally interchangeable with the yeast zir

A:Reference number: S53013

A:Accession: S53014

A&gt;Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-58 &lt;VIG&gt;

C:Superfamily: DNA-directed RNA polymerase chain ABC10 alpha

Query Match 8.1%; Score 26.5; DB 2; Length 58;  
Best Local Similarity 33.3%; Pred. No. 2.7e+04;  
Matches 7; Conservative 4; Mismatches 5; Indels 5; Gaps 2;

QY 22 KDLNPSR----YLC-KCPNE 37

DB 5 KDVPQPKQPMIYICGECHE 25

RESULT 30  
C83918  
hypothetical protein BH2147 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004

C:Accession: C83918

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiran  
Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: C83918

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-58 &lt;STO&gt;

A:Cross-references: UNIPROT:Q9KAY9; GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BAB058

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH2147

Query Match 8.1%; Score 26.5; DB 2; Length 58;  
Best Local Similarity 33.3%; Pred. No. 2.7e+04;  
Matches 9; Conservative 4; Mismatches 7; Indels 7; Gaps 2;

QY 6 CAEKTFCVNGGECFMVKDLSNPSRYLC 32

DB 8 CAT---CVH----FEVEKVNNGMVYRC 27

Search completed: April 14, 2005, 08:44:54

Job time : 49 secs

\_\_\_\_\_

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2005, 08:44:18 ; Search time 131 Seconds

(without alignments)

147.156 Million cell updates/sec

Title: US-09-980-672-2

Perfect score: 327

Sequence: 1 SHLVKCAKFTCVNGBCFM.....TGDRCQNVWMAFFYKABEY 58

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1421835 seqs, 332370683 residues

Total number of hits satisfying chosen parameters: 4758

Minimum DB seq length: 58

Maximum DB seq length: 58

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	67.5	20.6	58	9	US-09-934-706-3
2	67.5	20.6	58	15	US-10-344-634-12
3	53	16.2	58	16	US-10-311-828A-15
4	50	15.3	58	15	US-10-424-599-172436
5	48.5	14.8	58	14	US-10-231-778-231
6	46	14.1	58	9	US-09-981-649A-11
7	46	14.1	58	14	US-10-124-986-11
8	46	14.1	58	14	US-10-136-227A-11
9	46	14.1	58	14	US-10-112-881-11
10	46	14.1	58	15	US-10-399-123-11
11	44.5	13.6	58	13	US-10-044-359-2
12	42	12.8	58	15	US-10-424-599-160572
13	42	12.8	58	15	US-10-424-599-200587

41.5	12.7	58	16	US-10-767-701-58917	Sequence 58917, A
41	12.5	58	15	US-10-424-599-174130	Sequence 174130,
39	11.9	58	16	US-10-437-963-129880	Sequence 129880,
38.5	11.8	58	14	US-10-087-887-80	Sequence 80, Appl
38	11.6	58	15	US-10-424-599-237614	Sequence 237614,
37.5	11.5	58	9	US-09-764-887-251	Sequence 251, Appl
37.5	11.5	58	13	US-10-044-359-18	Sequence 18, Appl
37.5	11.5	58	14	US-10-073-961-251	Sequence 251, Appl
37.5	11.5	58	15	US-10-424-599-237766	Sequence 237766,
37	11.3	58	15	US-10-424-599-154581	Sequence 154581,
37	11.3	58	15	US-10-424-599-267478	Sequence 267478,
36	11.0	58	11	US-09-864-408A-3516	Sequence 3516, Ap
36	11.0	58	15	US-10-424-599-150646	Sequence 150646,
36	11.0	58	15	US-10-424-599-228516	Sequence 228516,
35.5	10.9	58	9	US-09-864-761-34904	Sequence 34904, A
35.5	10.9	58	15	US-10-424-599-227992	Sequence 227992,
35.5	10.9	58	16	US-10-437-963-137085	Sequence 137085,
35.5	10.9	58	16	US-10-437-963-157025	Sequence 157025,
35	10.7	58	9	US-09-829-155C-3	Sequence 3, Appli
35	10.7	58	15	US-10-424-599-178654	Sequence 178654,
35	10.7	58	16	US-10-437-963-186375	Sequence 186375,
34.5	10.6	58	9	US-09-864-761-37619	Sequence 37619, A
34.5	10.6	58	15	US-10-424-599-210806	Sequence 210806,
34.5	10.6	58	16	US-10-437-963-153395	Sequence 153395,
34	10.4	58	9	US-09-925-301-991	Sequence 991, App
34	10.4	58	14	US-10-097-111-404	Sequence 404, App
34	10.4	58	14	US-10-167-351-68	Sequence 68, Appl
34	10.4	58	14	US-10-038-722-62	Sequence 62, Appl
34	10.4	58	15	US-10-291-265-355	Sequence 355, App
34	10.4	58	15	US-10-291-265-827	Sequence 827, App
34	10.4	58	15	US-10-360-101-90	Sequence 90, Appl
34	10.4	58	15	US-10-424-599-275530	Sequence 275530,
34	10.4	58	15	US-10-456-986A-36	Sequence 36, Appl
34	10.4	58	16	US-10-437-963-159330	Sequence 159330,
34	10.4	58	16	US-10-361-997-24	Sequence 24, Appl
34	10.4	58	16	US-10-767-701-47537	Sequence 47537, A
33.5	10.2	58	9	US-09-781-988-56	Sequence 56, Appl
33.5	10.2	58	10	US-09-893-878-56	Sequence 56, Appl
33.5	10.2	58	10	US-09-896-095-56	Sequence 56, Appl
33.5	10.2	58	15	US-10-126-685-56	Sequence 56, Appl
33.5	10.2	58	15	US-10-127-028-56	Sequence 56, Appl
33.5	10.2	58	15	US-10-126-544-56	Sequence 56, Appl
33.5	10.2	58	15	US-10-425-114-50730	Sequence 50730, A
33	10.1	58	9	US-09-730-379B-2	Sequence 2, Appli
33	10.1	58	9	US-09-730-379B-12	Sequence 12, Appl
33	10.1	58	9	US-09-864-761-36918	Sequence 36918, A
33	10.1	58	9	US-09-864-761-40910	Sequence 40910, A
33	10.1	58	9	US-09-779-054-21	Sequence 21, Appl
33	10.1	58	15	US-10-467-020-6	Sequence 6, Appli
33	10.1	58	16	US-10-437-963-149342	Sequence 149342,
33	10.1	58	16	US-10-437-963-197828	Sequence 197828,
33	10.1	58	16	US-10-437-963-175111	Sequence 175111,
33	10.1	58	16	US-10-437-963-175111	Sequence 175111,
32.5	9.9	58	14	US-10-029-386-29573	Sequence 29573, A
32.5	9.9	58	15	US-10-424-599-216899	Sequence 216899,
32.5	9.9	58	15	US-10-424-599-253211	Sequence 253211,
32.5	9.9	58	16	US-10-437-963-109799	Sequence 109799,
32	9.8	58	9	US-09-764-878-191	Sequence 191, App
32	9.8	58	9	US-09-731-449-17	Sequence 17, Appl
32	9.8	58	9	US-09-874-056-5	Sequence 5, Appli
32	9.8	58	9	US-09-874-056-7	Sequence 7, Appli
32	9.8	58	14	US-10-079-854-191	Sequence 191, App
32	9.8	58	14	US-10-254-426-17	Sequence 17, Appl
32	9.8	58	15	US-10-424-599-152492	Sequence 152492,
32	9.8	58	15	US-10-424-599-175111	Sequence 175111,
32	9.8	58	15	US-10-424-599-200518	Sequence 200518,
32	9.8	58	15	US-10-424-599-228633	Sequence 228633,
32	9.8	58	16	US-10-437-963-112370	Sequence 112370,
32	9.8	58	16	US-10-437-963-184397	Sequence 184397,
32	9.8	58	16	US-10-178-213-234	Sequence 234, App
31.5	9.6	58	14	US-10-054-967-36	Sequence 36, Appl
31.5	9.6	58	14	US-10-054-967-81	Sequence 81, Appl
31.5	9.6	58	14	US-10-106-698-5599	Sequence 5599, Ap
31.5	9.6	58	14	US-10-231-778-229	Sequence 229, App









```
RESULT 10
US-10-399-123-11
; Sequence 11, Application US/10399123
; Publication No. US2004005908A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
; FILE REFERENCE: 28110/37665
; CURRENT APPLICATION NUMBER: US/10/399,123
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Mammalian
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(58)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-399-123-11

Query Match      14.1%; Score 46; DB 15; Length 58;
Best Local Similarity 28.2%; Pred. No. 5e+02;
Matches 11; Conservative 8; Mismatches 12; Indels 8; Gaps 3;

Qy 11 FCVNGGCFMVKDLSNPSRYLCKCPNEF--TGDR--CQ 44
Db 13 YCLHGVCMYTEAL---DRYACNCVVGYIXXGGRXXCQ 48

RESULT 11
US-10-044-359-2
; Sequence 2, Application US/10044359
; Publication No. US20020160454A1
; GENERAL INFORMATION:
; APPLICANT: Herriman, Rafael
; APPLICANT: Wong, James F.
; APPLICANT: Lee, Jian-Ming
; TITLE OF INVENTION: SCORPION TOXINS
; FILE REFERENCE: BB1367 US NA
; CURRENT APPLICATION NUMBER: US/10/044,359
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: 09/599,416
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,227
; PRIOR FILING DATE: 1999-06-22
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Hottentotta judaica
US-10-044-359-2

Query Match      13.6%; Score 44.5; DB 13; Length 58;
Best Local Similarity 38.7%; Pred. No. 7.6e+02;
Matches 12; Conservative 1; Mismatches 7; Indels 11; Gaps 2;

Qy 4 VKCAEKTFCVNGGCFMVKDLSNPSRYLCKC 34
Db 37 VKCAAMGF--NSGKCINSK-----CKC 56

RESULT 12
```

```
US-10-424-599-160572
; Sequence 160572, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 160572
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_116014C.1.pep
US-10-424-599-160572

Query Match      12.8%; Score 42; DB 15; Length 58;
Best Local Similarity 32.5%; Pred. No. 1.5e+03;
Matches 13; Conservative 5; Mismatches 16; Indels 6; Gaps 2;

Qy 2 HLVKCAEKTFCVN--GGECFMVKDLSNPSRYLCKCPNEF 38
Db 1 HIMETREVFQLNRTGGGFLMV--LGPPNYFPQSPNTF 37

RESULT 13
US-10-424-599-200587
; Sequence 200587, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 200587
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_23156C.1.pep
US-10-424-599-200587

Query Match      12.8%; Score 42; DB 15; Length 58;
Best Local Similarity 30.6%; Pred. No. 1.5e+03;
Matches 11; Conservative 8; Mismatches 11; Indels 6; Gaps 2;

Qy 1 SHLVKCAEKTFCVNGGCFMVKDLSNPSRYLCKCPN 36
Db 26 ARLHKEAESTWCI----CHI--NSNNPLKVLVLSCKN 55

RESULT 14
US-10-767-701-58917
; Sequence 58917, Application US/10767701
; Publication No. US2004017684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 58917
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_23156C.1.pep
US-10-767-701-58917

Query Match      12.8%; Score 42; DB 15; Length 58;
Best Local Similarity 30.6%; Pred. No. 1.5e+03;
Matches 11; Conservative 8; Mismatches 11; Indels 6; Gaps 2;

Qy 1 SHLVKCAEKTFCVNGGCFMVKDLSNPSRYLCKCPN 36
Db 26 ARLHKEAESTWCI----CHI--NSNNPLKVLVLSCKN 55
```

```

; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 58917
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; -OTHER INFORMATION: Clone ID: 6858026.pcp
US-10-767-701-58917

```

Query Match 12.7%; Score 41.5; DB 16; Length 58;  
Best Local Similarity 26.8%; Pred. No. 1.7e+03;  
Matches 11; Conservative 6; Mismatches 17; Indels

**Qy** 1 SHLVKCAEKTFCVNGGE-----CFMVKDLSNP SRYLCKCPN 36  
:  
||| : : | : : | :  
**Dd** 7 NHYAKCGEFVYSILAGKKIOOVCLIIFD--NSINLLLFCON 45

```

RESULT 15
US-10-424-599-174130
; Sequence 174130, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molec
; TITLE OF INVENTION: Plants and Uses There
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 174130
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; - OTHER INFORMATION: Clone ID: PAT MRT3847
US-10-424-599-174130

```

Query Match 12.5%; Score 41; DB\_15; Length 58;  
Best Local Similarity 36.4%; Pred. No. 2e+03;  
Matches 8; Conservative 4; Mismatches 6; Indels

Qy 3 LVKCAEKTFCVNGGE----CFM 20  
: | | | | : : : | :  
pb 8 VVHCAGPTLCIGVGVOVSPCEFL 29

```

RESULT 16
US-10-437-963-129880
; Sequence 129880, Application US/10437963
; Publication No. US20040123343A1
;
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbasuk, Brad
; APPLICANT: Li, Ping
;
; TITLE OF INVENTION: Rice Nucleic Acid Molecules
;
; FILE REFERENCE: 38-21(53221)B
;
; CURRENT APPLICATION NUMBER: US/10/437,963
;
; CURRENT FILING DATE: 2003-05-14
;
; NUMBER OF SEQ ID NOS: 204966
;
; SEQ ID NO 129880

```

```

; LENGTH: 58
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(58)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_32094C.1.pep
US-10-437-963-129880

```

```

Query Match      11.9%; Score 39; DB 16; Length 58;
Best Local Similarity 26.7%; Pred.No. 3.4e+03;
Matches 8; Conservative 6; Mismatches 16; Indels

QY      6 CAEKTCFVNGGECFMVKDLSNPSRYLKCP 35
Db      6 CADTSLCVKATATCYGVGLHKAENHHKXTP 35

```

```

RESULT 17
US-10-087-887-80
; Sequence 80, Application US/10087887
; Publication No. US20030198957A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Conley, Pamela B.
; APPLICANT: Yang, Ruey-Bing
; APPLICANT: Hart, Matthew
; APPLICANT: Tomlinson, James E.
; APPLICANT: Topper, James N.
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Leach, Martin D.
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Komuves, Laszlo
; APPLICANT: Padigaru, Muralidhara
; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-285
; CURRENT APPLICATION NUMBER: US/10/087,887
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/273,049
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/279,883
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/277,791
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/281,248
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/282,864
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/282,537
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: 60/282,867
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 104
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 80
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-087-887-80

```

Query Match	11.8%;	Score 38.5;	DB 14;	Length 58;
Best Local Similarity	29.5%;	Pred. No. 3.9e+03;		
Matches 13; Conservative	6;	Mismatches 8;	Indels 17;	

**Qy**            5 KCAE-----KTF-----VNGGECFMVKDSLNPRLCKC 34  
               :|||       |||       |||       |       |       |  
**Db**            12 ECTAEGRHWIMKHFCFECETVLGGORYIMKE-GRP--YCCHC 52

RESULT 18  
US-10-424-599-237614

```

; Sequence 237614, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 237614
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_56590C.1.pap
US-10-424-599-237614

Query Match      11.6%; Score 38; DB 15; Length 58;
Best Local Similarity 54.5%; Pred. No. 4.4e+03;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy      26 NPSRYLCKCPN 36
Db      37 NPNFYIKTCPN 47

RESULT 19
US-09-764-887-251
; Sequence 251, Application US/09764887
; Patent No. US20020042096A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PALL3
; CURRENT APPLICATION NUMBER: US/09/764,887
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 658
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 251
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-887-251

Query Match      11.5%; Score 37.5; DB 9; Length 58;
Best Local Similarity 77.8%; Pred. No. 5.1e+03;
Matches 7; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Qy      30 YLCKCPNEF 38
Db      13 HLC-CPNEF 20

RESULT 20
US-10-044-359-18
; Sequence 18, Application US/10044359
; Publication No. US20020160454A1
; GENERAL INFORMATION:
; APPLICANT: Heriman, Rafael
; APPLICANT: Wong, James F.
; APPLICANT: Lee, Jian-Ming
; TITLE OF INVENTION: SCORPION TOXINS
; FILE REFERENCE: BBI367 US NA
; CURRENT APPLICATION NUMBER: US/10/044,359
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: 09/599,416
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,227

```

```

; PRIOR FILING DATE: 1999-06-22
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Hottentotta judaica
US-10-044-359-18

Query Match      11.5%; Score 37.5; DB 13; Length 58;
Best Local Similarity 27.8%; Pred. No. 5.1e+03;
Matches 10; Conservative 10; Mismatches 9; Indels 7; Gaps 3;

Qy      12 CVNGGECFMV-KDLSNPSRYLCKCPNEFTGDRCONY 46
Db      29 CISSQECWIAKCKVT--GRFGKCKNK----QCRCY 58

RESULT 21
US-10-073-961-251
; Sequence 251, Application US/10073961
; Publication No. US20030077602A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PALL3C1
; CURRENT APPLICATION NUMBER: US/10/073,961
; CURRENT FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: 09/764,887
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/225,757
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/226,868
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/216,647
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,267
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/216,880
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,270
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/251,869
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/235,834
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/234,274
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/234,223
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/228,924
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/224,518
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,369

```

;  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/224,519  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,964  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/241,809  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/249,299  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/236,327  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/241,785  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/244,617  
; PRIOR FILING DATE: 2000-11-01  
; PRIOR APPLICATION NUMBER: 60/225,268  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/236,368  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/251,856  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/251,868  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/229,344  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/234,997  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: 60/229,343  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,345  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,287  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,513  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: 60/231,413  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/229,509  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: 60/236,367  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/237,039  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,038  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/236,370  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/236,802  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,037  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,040  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/240,960  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/239,935  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: 60/239,937  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: 60/241,787  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/246,474  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/246,532  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/249,216  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,210  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/226,681  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/225,759  
; PRIOR FILING DATE: 2000-08-14

;  
; PRIOR APPLICATION NUMBER: 60/225,213  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/227,182  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/225,214  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/235,836  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/230,438  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/215,135  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: 60/225,266  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/249,218  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,208  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,213  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,212  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,207  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,245  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,244  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,217  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,211  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,215  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,264  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,214  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,297  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/232,400  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/231,242  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/232,081  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/232,080  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/231,414  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/231,244  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/233,064  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/233,063  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/232,397  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/232,399  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/232,401  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/241,808  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,826  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,786  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,221  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/246,475  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/231,243

DB 18 PSKAVCACPH 27

```
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 150646
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_107058C.1.pep
US-10-424-599-150646

Query Match      11.0%; Score 36; DB 15; Length 58;
Best Local Similarity 32.1%; Pred. No. 7.6e+03;
Matches 9; Conservative 4; Mismatches 11; Indels 4; Gaps 1;

Qy 5 KCAEKTFCVNGGECFMVKDLSNPGRYLC 32
Db 35 KVFSSSFCIN---FPLKHLKSSSFEFC 58

RESULT 27
US-10-424-599-228516
; Sequence 228516, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 228516
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_48378C.1.pep
US-10-424-599-228516

Query Match      11.0%; Score 36; DB 15; Length 58;
Best Local Similarity 56.2%; Pred. No. 7.6e+03;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 21 VKDLSNPGRYLCPCPN 36
Db 28 VKNISYRSRYLDLEPN 43

RESULT 28
US-09-864-761-34904
; Sequence 34904, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
```

```
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34904
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC011092.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.91
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.97
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.91
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EST_HUMAN HIT: BE256927.1, EVALUATE 2.00e-22
US-09-864-761-34904

Query Match      10.9%; Score 35.5; DB 9; Length 58;
Best Local Similarity 25.0%; Pred. No. 8.7e+03;
Matches 14; Conservative 5; Mismatches 12; Indels 25; Gaps 3;

Qy 11 FCVNGG-----ECFMVKDLS--NPSRYLCKPCNEFTGDRCONVMAFYKAE 55
Db 1 FCOPGGWQLSRERKQPTFFVVLTDIDSDRHYCSC-----LTFYEAE 42

RESULT 29
US-10-424-599-227992
; Sequence 227992, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
```

```

; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 227992
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_47904C.1.pap
US-10-424-599-227992

Query Match      10.9%; Score 35.5; DB 15; Length 58;
Best Local Similarity 28.9%; Pred. No. 8.7e+03;
Matches 11; Conservative 4; Mismatches 12; Indels 11; Gaps 2;

QY      8 EKTFCVNGGECFMVKD----LSNPSRYLC-----KC 34
DB      12 QSTFCVHRTACTFLPRFAPVLNWPGRFAFCVTFKKIKC 49

RESULT 30
US-10-437-963-137085
; Sequence 137085, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: Thomas J.
; APPLICANT: La Rosa, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 137085
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_38601C.1.pap
US-10-437-963-137085

Query Match      10.9%; Score 35.5; DB 16; Length 58;
Best Local Similarity 23.1%; Pred. No. 8.7e+03;
Matches 9; Conservative 6; Mismatches 17; Indels 7; Gaps 1;

QY      8 EKTFCVNGGECFMVKDLSNPSRYLCKPNEFTGDRCONY 46
DB      2 DQSIDGLYSCDQKREKACNPTF-----AHTSESLNY 33

Search completed: April 14, 2005, 09:00:13
Job time : 134 secs

```

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2005, 08:35:17 ; Search time 43 Seconds

(without alignments)

100.689 Million cell updates/sec

Title: US-09-980-672-2

Perfect score: 327

Sequence: 1 SHLVKAEKTFVNGGECFM.....TGDRCONVMVFYKAEELY 58

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 2483

Minimum DB seq length: 58

Maximum DB seq length: 58

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Issued Patents:AA:  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	14.7	58	5189019-8	Patent No. 5189019
2	48	14.7	58	5189019-8	Patent No. 5189019
3	46	14.1	58	US-09-249-697A-11	Sequence 11, Appl
4	46	14.1	58	US-09-363-316B-11	Sequence 11, Appl
5	46	14.1	58	US-09-513-999C-7434	Sequence 7434, Ap
6	46	14.1	58	US-10-136-227A-11	Sequence 11, Appl
7	46	14.1	58	US-09-981-649A-11	Sequence 11, Appl
8	44.5	13.6	58	US-10-044-359-2	Sequence 2, Appl
9	39.5	12.1	58	US-08-261-206A-3	Sequence 3, Appl
10	37.5	11.5	58	US-09-270-767-57784	Sequence 18, Appl
11	37.5	11.5	58	US-10-044-359-18	Sequence 14, Appl
12	36	11.0	58	US-08-904-446A-14	Sequence 1294, Ap
13	35	10.7	58	US-09-471-276-1294	Sequence 13, Appl
14	34.5	10.6	58	US-08-086-630C-13	Sequence 16, Appl
15	34.5	10.6	58	US-08-086-630C-16	Sequence 20, Appl
16	34.5	10.6	58	US-08-086-630C-20	Sequence 21, Appl
17	34.5	10.6	58	US-08-086-630C-21	Sequence 22, Appl
18	34.5	10.6	58	US-08-086-630C-22	Sequence 23, Appl
19	34.5	10.6	58	US-08-086-630C-23	Sequence 24, Appl
20	34.5	10.6	58	US-08-086-630C-24	Sequence 25, Appl
21	34.5	10.6	58	US-08-086-630C-25	Sequence 26, Appl
22	34.5	10.6	58	US-08-086-630C-26	Sequence 112, App
23	34.5	10.6	58	US-08-086-630C-112	Sequence 114, App
24	34.5	10.6	58	US-08-086-630C-114	Sequence 118, App
25	34.5	10.6	58	US-08-086-630C-118	Sequence 119, App
26	34.5	10.6	58	US-08-086-630C-119	Sequence 121, App
27	34.5	10.6	58	US-08-086-630C-121	

28	34.5	10.6	58	1	US-08-086-630C-145	Sequence 145, App
29	34.5	10.6	58	1	US-08-086-630C-152	Sequence 152, App
30	34.5	10.6	58	1	US-08-086-630C-153	Sequence 153, App
31	34.5	10.6	58	1	US-08-086-630C-154	Sequence 154, App
32	34.5	10.6	58	1	US-08-086-630C-155	Sequence 155, App
33	34.5	10.6	58	1	US-08-086-630C-157	Sequence 157, App
34	34.5	10.6	58	1	US-08-086-630C-189	Sequence 189, App
35	34.5	10.6	58	1	US-08-086-630C-192	Sequence 192, App
36	34.5	10.6	58	1	US-08-086-630C-196	Sequence 196, App
37	34.5	10.6	58	1	US-08-086-630C-197	Sequence 197, App
38	34.5	10.6	58	1	US-08-086-630C-198	Sequence 198, App
39	34.5	10.6	58	1	US-08-086-630C-199	Sequence 199, App
40	34.5	10.6	58	1	US-08-086-630C-200	Sequence 200, App
41	34.5	10.6	58	1	US-08-086-630C-201	Sequence 201, App
42	34.5	10.6	58	1	US-08-086-630C-202	Sequence 202, App
43	34.5	10.6	58	1	US-08-086-328C-10	Sequence 10, Appl
44	34.5	10.6	58	1	US-08-086-328C-16	Sequence 16, Appl
45	34.5	10.6	58	1	US-08-086-328C-119	Sequence 119, App
46	34.5	10.6	58	1	US-08-086-328C-120	Sequence 120, App
47	34.5	10.6	58	1	US-08-086-328C-121	Sequence 121, App
48	34.5	10.6	58	1	US-08-086-328C-122	Sequence 122, App
49	34.5	10.6	58	1	US-08-086-328C-125	Sequence 125, App
50	34.5	10.6	58	1	US-08-086-328C-126	Sequence 126, App
51	34.5	10.6	58	1	US-08-086-328C-140	Sequence 140, App
52	34.5	10.6	58	1	US-08-086-328C-146	Sequence 146, App
53	34.5	10.6	58	1	US-08-086-328C-189	Sequence 189, App
54	34.5	10.6	58	1	US-08-086-328C-195	Sequence 195, App
55	34	10.4	58	1	US-08-058-699-3	Sequence 3, Appl
56	34	10.4	58	1	US-08-384-489-16	Sequence 16, Appl
57	34	10.4	58	1	US-08-358-160-79	Sequence 79, Appl
58	34	10.4	58	1	US-08-463-155A-40	Sequence 40, Appl
59	34	10.4	58	1	US-08-463-432B-40	Sequence 40, Appl
60	34	10.4	58	1	US-08-676-125A-37	Sequence 37, Appl
61	34	10.4	58	1	US-08-206-310A-40	Sequence 40, Appl
62	34	10.4	58	2	US-08-398-010A-40	Sequence 40, Appl
63	34	10.4	58	2	US-08-398-628A-40	Sequence 40, Appl
64	34	10.4	58	2	US-08-399-115A-40	Sequence 40, Appl
65	34	10.4	58	2	US-09-136-012A-37	Sequence 37, Appl
66	34	10.4	58	3	US-08-676-124-68	Sequence 68, Appl
67	34	10.4	58	3	US-09-414-878-68	Sequence 68, Appl
68	34	10.4	58	3	US-09-240-136-68	Sequence 68, Appl
69	34	10.4	58	3	US-09-369-494-4	Sequence 4, Appl
70	34	10.4	58	3	US-09-358-569D-4	Sequence 4, Appl
71	34	10.4	58	4	US-09-638-770A-68	Sequence 68, Appl
72	34	10.4	58	4	US-09-569-670-4	Sequence 4, Appl
73	33.5	10.2	58	1	US-07-664-989B-56	Sequence 56, Appl
74	33.5	10.2	58	1	US-08-358-160-33	Sequence 33, Appl
75	32.5	9.9	58	1	US-08-086-630C-10	Sequence 10, Appl
76	32.5	9.9	58	1	US-08-086-630C-11	Sequence 11, Appl
77	32.5	9.9	58	1	US-08-086-630C-12	Sequence 12, Appl
78	32.5	9.9	58	1	US-08-086-630C-14	Sequence 14, Appl
79	32.5	9.9	58	1	US-08-086-630C-17	Sequence 17, Appl
80	32.5	9.9	58	1	US-08-086-630C-18	Sequence 18, Appl
81	32.5	9.9	58	1	US-08-086-630C-92	Sequence 92, Appl
82	32.5	9.9	58	1	US-08-086-630C-109	Sequence 109, App
83	32.5	9.9	58	1	US-08-086-630C-110	Sequence 110, App
84	32.5	9.9	58	1	US-08-086-630C-111	Sequence 111, App
85	32.5	9.9	58	1	US-08-086-630C-113	Sequence 113, App
86	32.5	9.9	58	1	US-08-086-630C-115	Sequence 115, App
87	32.5	9.9	58	1	US-08-086-630C-116	Sequence 116, App
88	32.5	9.9	58	1	US-08-086-630C-120	Sequence 120, App
89	32.5	9.9	58	1	US-08-086-630C-143	Sequence 143, App
90	32.5	9.9	58	1	US-08-086-630C-144	Sequence 144, App
91	32.5	9.9	58	1	US-08-086-630C-146	Sequence 146, App
92	32.5	9.9	58	1	US-08-086-630C-149	Sequence 149, App
93	32.5	9.9	58	1	US-08-086-630C-150	Sequence 150, App
94	32.5	9.9	58	1	US-08-086-630C-156	Sequence 156, App
95	32.5	9.9	58	1	US-08-086-630C-186	Sequence 186, App
96	32.5	9.9	58	1	US-08-086-630C-187	Sequence 187, App
97	32.5	9.9	58	1	US-08-086-630C-188	Sequence 188, App
98	32.5	9.9	58	1	US-08-086-630C-190	Sequence 190, App
99	32.5	9.9	58	1	US-08-086-630C-193	Sequence 193, App
100	32.5	9.9	58	1	US-08-086-630C-194	Sequence 194, App



```
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 25
; OTHER INFORMATION: Xaa=Leu or Val
US-09-513-999C-7434

Query Match      14.1%; Score 46; DB 4; Length 58;
Best Local Similarity 36.4%; Pred. No. 2e+02;
Matches 12; Conservative 4; Mismatches 9; Indels 8; Gaps 3;

Oy 6 CAEKTFCVNGGECF--MVKDLSNPSRYLCKCPN 36
Db 10 CRVRVAC-----CFANMVFNKXNPKSFLC-CPH 36

RESULT 6
US-10-136-227A-11
; Sequence 11, Application US/10136227A
; Patent No. 6800443
; GENERAL INFORMATION:
; APPLICANT: Ford et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
; FILE REFERENCE: 28110/38407
; CURRENT APPLICATION NUMBER: US/10/136,227A
; CURRENT FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 09/981,649
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Mammalian
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (39)
; OTHER INFORMATION: Xaa = Any Amino Acid
; NAME/KEY: misc_feature
; LOCATION: (40)
; OTHER INFORMATION: Xaa = Any Amino Acid
; NAME/KEY: misc_feature
; LOCATION: (41)
; OTHER INFORMATION: Xaa = Any Amino Acid
; NAME/KEY: misc_feature
; LOCATION: (45)
; OTHER INFORMATION: Xaa = Any Amino Acid
; NAME/KEY: misc_feature
; LOCATION: (46)
; OTHER INFORMATION: Xaa = Any Amino Acid
; US-09-981-649A-11
; SEQ ID NO 11
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Mammalian
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (39)
; OTHER INFORMATION: Xaa = Any Amino Acid
; NAME/KEY: misc_feature
; LOCATION: (40)
; OTHER INFORMATION: Xaa = Any Amino Acid
; NAME/KEY: misc_feature
; LOCATION: (41)
; OTHER INFORMATION: Xaa = Any Amino Acid
; NAME/KEY: misc_feature
; LOCATION: (45)
; OTHER INFORMATION: Xaa = Any Amino Acid
; NAME/KEY: misc_feature
; LOCATION: (46)
; OTHER INFORMATION: Xaa = Any Amino Acid
; US-10-136-227A-11

Query Match      14.1%; Score 46; DB 4; Length 58;
Best Local Similarity 28.2%; Pred. No. 2e+02;
Matches 11; Conservative 8; Mismatches 12; Indels 8; Gaps 3;

Oy 11 FCVNGGECFMVKDLSNPSRYLCKCPNEF---TGDR--CQ 44
Db 13 YCLHDGVCWYIEAL---DKYACNCVVGVIYXXGEXXCQ 48

RESULT 7
US-09-981-649A-11
; Sequence 11, Application US/09981649A
; Patent No. 6740743
; GENERAL INFORMATION:
; APPLICANT: Herrman, Rafael
; APPLICANT: Wong, James F.
; APPLICANT: Lee, Jian-Ming
; TITLE OF INVENTION: SCORPION TOXINS
; FILE REFERENCE: BB1367 US NA
; CURRENT APPLICATION NUMBER: US/10/044,359
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: 09/599,416
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,227
; PRIOR FILING DATE: 1999-06-22
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Hottentotta judaica
; US-10-044-359-2

Query Match      13.6%; Score 44.5; DB 4; Length 58;
Best Local Similarity 38.7%; Pred. No. 3e+02;
Matches 12; Conservative 1; Mismatches 7; Indels 11; Gaps 2;
```



STREET: 1215 Jefferson Davis Highway, Suite 309  
CITY: Arlington  
STATE: Virginia  
COUNTRY: United States of America  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/904,446A  
FILING DATE: 31-JUL-1997  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: GB 9616105.4  
FILING DATE: 31-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: RALPH A. DOWELL  
REGISTRATION NUMBER: 26868  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-2555  
TELEFAX: (703) 415-2559  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 58 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..58  
OTHER INFORMATION: /note= "Identified neurotrophin  
OTHER INFORMATION: binding site of human p75NTR, residues 39-96"  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..2  
OTHER INFORMATION: /note= "cysteine 39 forms a  
OTHER INFORMATION: disulfide bond with cysteine 55"  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 17..18  
OTHER INFORMATION: /note= "cysteine 55 forms a  
OTHER INFORMATION: disulfide bond with cysteine 39"  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 20..21  
OTHER INFORMATION: /note= "cysteine 58 forms a  
OTHER INFORMATION: disulfide bond with cysteine 71"  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 33..34  
OTHER INFORMATION: /note= "cysteine 71 forms a  
OTHER INFORMATION: disulfide bond with cysteine 58"  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 23..24  
OTHER INFORMATION: /note= "cysteine 61 forms a  
OTHER INFORMATION: disulfide bond with cysteine 79"  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 41..42  
OTHER INFORMATION: /note= "cysteine 79 forms a  
OTHER INFORMATION: disulfide bond with cysteine 61"  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 43..44  
OTHER INFORMATION: /note= "cysteine 81 forms a  
OTHER INFORMATION: disulfide bond with cysteine 94"  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 56..57  
OTHER INFORMATION: /note= "cysteine 94 forms a

OTHER INFORMATION: disulfide bond with cysteine 81"  
US-08-904-446A-14  
Query Match 11.0%; Score 36; DB 3; Length 58;  
Best Local Similarity 28.3%; Pred. No. 2.9e+03;  
Matches 15; Conservative 8; Mismatches 18; Indels 12; Gaps 4;  
QY 1 SHLVKCAKTECVNGGECFMVKDLNPN-----SRYLCKC-----PNEFTGDRQ 44  
DB 8 SDVVSATEP--CKPCTECVGLQMSAPCVCEADDAVCRCAYGYQDETGG-RCE 57

## RESULT 13

US-09-471-276-1294  
Sequence 1294, Application US/09471276  
Patent No. 6822072  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, J.B.  
APPLICANT: Duclert A.  
APPLICANT: Giordano, J.Y.  
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
Patent No. 6822072  
FILE REFERENCE: GENSET.025CPI  
CURRENT APPLICATION NUMBER: US/09/471,276  
CURRENT FILING DATE: 1999-12-21  
EARLIER APPLICATION NUMBER: 09/057,719  
EARLIER FILING DATE: 1998-04-09  
EARLIER APPLICATION NUMBER: 09/069,047  
EARLIER FILING DATE: 1998-04-28  
EARLIER APPLICATION NUMBER: PCT/IB99/00712  
EARLIER FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 1622  
SOFTWARE: Patent.pm  
SEQ ID NO 1294  
LENGTH: 58  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SIGNAL  
LOCATION: -21..-1  
US-09-471-276-1294

Query Match 10.7%; Score 35; DB 4; Length 58;  
Best Local Similarity 31.6%; Pred. No. 3.8e+03;  
Matches 6; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 SHLVKCAKTECVNGGECF 19  
DB 2 SPLQAVIRAVCLSGGSCW 20

## RESULT 14

US-08-086-630C-13  
Sequence 13, Application US/08086630C  
Patent No. 5747449  
GENERAL INFORMATION:  
APPLICANT: Ignace Lasters, Marc De Maeyer  
APPLICANT: and William Charles Ripka  
TITLE OF INVENTION: BOVINE PANCREATIC TRYPSIN INHIBITOR  
TITLE OF INVENTION: DERIVED INHIBITORS OF FACTOR Xa  
NUMBER OF SEQUENCES: 284  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 611 West Sixth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90017  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44Mb storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/086.630C
; FILING DATE: July 1, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 07/913,232
; FILING DATE: July 13, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 202/210
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-086-630C-13

Query Match 10.6%; Score 34.5; DB 1; Length 58;
Best Local Similarity 38.9%; Pred. No. 4.3e+03;
Matches 7; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Qy 38 FTGDRCONVWMAFYKAE 55
Db 10 YTG-P-CRAYIRYFYNK 26

RESULT 15
US-08-086-630C-16
; Sequence 16, Application US/08086630C
; Patent No. 5747449
; GENERAL INFORMATION:
; APPLICANT: Ignace Lasters, Marc De Maeyer
; APPLICANT: and William Charles Ripka
; TITLE OF INVENTION: BOVINE PANCREATIC TRYPSIN INHIBITOR
; TITLE OF INVENTION: DERIVED INHIBITORS OF FACTOR Xa
; NUMBER OF SEQUENCES: 284
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44Mb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/086.630C
; FILING DATE: July 1, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 07/913,232
; FILING DATE: July 13, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 202/210
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-086-630C-20

Query Match 10.6%; Score 34.5; DB 1; Length 58;
Best Local Similarity 38.9%; Pred. No. 4.3e+03;
Matches 7; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Qy 38 FTGDRCONVWMAFYKAE 55
Db 10 YTG-P-CRAYIRYFYNK 26
```

```
;
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-086-630C-16

Query Match 10.6%; Score 34.5; DB 1; Length 58;
Best Local Similarity 38.9%; Pred. No. 4.3e+03;
Matches 7; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Qy 38 FTGDRCONVWMAFYKAE 55
Db 10 YTG-ICRAYIRYFYNK 26

RESULT 16
US-08-086-630C-20
; Sequence 20, Application US/08086630C
; Patent No. 5747449
; GENERAL INFORMATION:
; APPLICANT: Ignace Lasters, Marc De Maeyer
; APPLICANT: and William Charles Ripka
; TITLE OF INVENTION: BOVINE PANCREATIC TRYPSIN INHIBITOR
; TITLE OF INVENTION: DERIVED INHIBITORS OF FACTOR Xa
; NUMBER OF SEQUENCES: 284
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44Mb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/086.630C
; FILING DATE: July 1, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 07/913,232
; FILING DATE: July 13, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 202/210
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-086-630C-20

Query Match 10.6%; Score 34.5; DB 1; Length 58;
Best Local Similarity 38.9%; Pred. No. 4.3e+03;
Matches 7; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Qy 38 FTGDRCONVWMAFYKAE 55
Db 10 YTG-ICRAYIRYFYNK 26
```



```
; APPLICATION NUMBER: 07/913,232
; FILING DATE: July 13, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 202/210
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-086-630C-23

Query Match 10.6%; Score 34.5; DB 1; Length 58;
Best Local Similarity 38.9%; Pred. No. 4.3e+03;
Matches 7; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Qy 38 FTGDRCONYVMASFYKAE 55
Db 10 YTGP-CRAYTRYFYNK 26

RESULT 20
US-08-086-630C-24
; Sequence 24, Application US/08086630C
; Patent No. 5747449
; GENERAL INFORMATION:
; APPLICANT: Ignace Lasters, Marc De Maeyer
; APPLICANT: and William Charles Ripka
; TITLE OF INVENTION: BOVINE PANCREATIC TRYPSIN INHIBITOR
; TITLE OF INVENTION: DERIVED INHIBITORS OF FACTOR Xa
; NUMBER OF SEQUENCES: 284
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44Mb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/086,630C
; FILING DATE: July 1, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 07/913,232
; FILING DATE: July 13, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 202/210
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-086-630C-24

Query Match 10.6%; Score 34.5; DB 1; Length 58;
Best Local Similarity 38.9%; Pred. No. 4.3e+03;
Matches 7; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Qy 38 FTGDRCONYVMASFYKAE 55
Db 10 YTGP-CRAYTRYFYNK 26

RESULT 21
US-08-086-630C-25
; Sequence 25, Application US/08086630C
; Patent No. 5747449
; GENERAL INFORMATION:
; APPLICANT: Ignace Lasters, Marc De Maeyer
; APPLICANT: and William Charles Ripka
; TITLE OF INVENTION: BOVINE PANCREATIC TRYPSIN INHIBITOR
; TITLE OF INVENTION: DERIVED INHIBITORS OF FACTOR Xa
; NUMBER OF SEQUENCES: 284
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44Mb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/086,630C
; FILING DATE: July 1, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 07/913,232
; FILING DATE: July 13, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 202/210
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-086-630C-25

Query Match 10.6%; Score 34.5; DB 1; Length 58;
Best Local Similarity 38.9%; Pred. No. 4.3e+03;
Matches 7; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Qy 38 FTGDRCONYVMASFYKAE 55
Db 10 YTGP-CRAYTRYFYNK 26

RESULT 22
US-08-086-630C-26
; Sequence 26, Application US/08086630C
; Patent No. 5747449
```

```
; MOLECULE TYPE: peptide
; US-08-086-630C-24

Query Match 10.6%; Score 34.5; DB 1; Length 58;
Best Local Similarity 38.9%; Pred. No. 4.3e+03;
Matches 7; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Qy 38 FTGDRCONYVMASFYKAE 55
Db 10 YTGP-CRAYTRYFYNK 26

RESULT 21
US-08-086-630C-25
; Sequence 25, Application US/08086630C
; Patent No. 5747449
; GENERAL INFORMATION:
; APPLICANT: Ignace Lasters, Marc De Maeyer
; APPLICANT: and William Charles Ripka
; TITLE OF INVENTION: BOVINE PANCREATIC TRYPSIN INHIBITOR
; TITLE OF INVENTION: DERIVED INHIBITORS OF FACTOR Xa
; NUMBER OF SEQUENCES: 284
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44Mb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/086,630C
; FILING DATE: July 1, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 07/913,232
; FILING DATE: July 13, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 202/210
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-086-630C-25

Query Match 10.6%; Score 34.5; DB 1; Length 58;
Best Local Similarity 38.9%; Pred. No. 4.3e+03;
Matches 7; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Qy 38 FTGDRCONYVMASFYKAE 55
Db 10 YTGP-CRAYTRYFYNK 26

RESULT 22
US-08-086-630C-26
; Sequence 26, Application US/08086630C
; Patent No. 5747449
```



```
; GENERAL INFORMATION:
; APPLICANT: Ignace Lasters, Marc De Maeyer
; APPLICANT: and William Charles Ripka
; TITLE OF INVENTION: BOVINE PANCREATIC TRYPSIN INHIBITOR
; TITLE OF INVENTION: DERIVED INHIBITORS OF FACTOR Xa
; NUMBER OF SEQUENCES: 284
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44Mb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/086,630C
; FILING DATE: July 1, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 07/913,232
; FILING DATE: July 13, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 202/210
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-086-630C-26

Query Match 10.6%; Score 34.5; DB 1; Length 58;
Best Local Similarity 38.9%; Pred. No. 4.3e+03;
Matches 7; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Qy 38 FTGDRCONVMSFYKAE 55
Db 10 YTGP-CRAYITRYFYNK 26

; RESULT 23
; US-08-086-630C-112
; Sequence 112, Application US/08086630C
; Patent No. 5747449
; GENERAL INFORMATION:
; APPLICANT: Ignace Lasters, Marc De Maeyer
; APPLICANT: and William Charles Ripka
; TITLE OF INVENTION: BOVINE PANCREATIC TRYPSIN INHIBITOR
; TITLE OF INVENTION: DERIVED INHIBITORS OF FACTOR Xa
; NUMBER OF SEQUENCES: 284
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44Mb storage
; COMPUTER: IBM Compatible
```

```
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/086,630C
; FILING DATE: July 1, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 07/913,232
; FILING DATE: July 13, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 202/210
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 112:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-086-630C-112

Query Match 10.6%; Score 34.5; DB 1; Length 58;
Best Local Similarity 38.9%; Pred. No. 4.3e+03;
Matches 7; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Qy 38 FTGDRCONVMSFYKAE 55
Db 10 YTGP-CRAYITRYFYNK 26

; RESULT 24
; US-08-086-630C-114
; Sequence 114, Application US/08086630C
; Patent No. 5747449
; GENERAL INFORMATION:
; APPLICANT: Ignace Lasters, Marc De Maeyer
; APPLICANT: and William Charles Ripka
; TITLE OF INVENTION: BOVINE PANCREATIC TRYPSIN INHIBITOR
; TITLE OF INVENTION: DERIVED INHIBITORS OF FACTOR Xa
; NUMBER OF SEQUENCES: 284
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44Mb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/086,630C
; FILING DATE: July 1, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 07/913,232
; FILING DATE: July 13, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 202/210
; TELECOMMUNICATION INFORMATION:
```

; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 114:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 58 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-086-630C-114

Query Match 10.6%; Score 34.5; DB 1; Length 58;  
Best Local Similarity 38.9%; Pred. No. 4.3e+03;  
Matches 7; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Qy 38 FTGDRCONVWASFYKAE 55  
Db 10 YTG-ICRAYIRYFYNK 26

## RESULT 25

US-08-086-630C-118  
; Sequence 118, Application US/08086630C  
; Patent No. 5747449  
; GENERAL INFORMATION:  
; APPLICANT: Ignace Lasters, Marc De Maeyer  
; APPLICANT: and William Charles Ripka  
; TITLE OF INVENTION: BOVINE PANCREATIC TRYPSIN INHIBITOR  
; TITLE OF INVENTION: DERIVED INHIBITORS OF FACTOR Xa  
; NUMBER OF SEQUENCES: 284  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 611 West Sixth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90017

COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44Mb storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/086.630C  
; FILING DATE: July 1, 1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA: including application  
; PRIOR APPLICATION DATA: described below:  
; APPLICATION NUMBER: 07/913,232  
; FILING DATE: July 13, 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 202/210  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 118:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 58 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-086-630C-118

Query Match 10.6%; Score 34.5; DB 1; Length 58;  
Best Local Similarity 38.9%; Pred. No. 4.3e+03;  
Matches 7; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Qy 38 FTGDRCONVWASFYKAE 55  
Db 10 YTG-ICRAYIRYFYNK 26

## RESULT 26

US-08-086-630C-119  
; Sequence 119, Application US/08086630C  
; Patent No. 5747449  
; GENERAL INFORMATION:  
; APPLICANT: Ignace Lasters, Marc De Maeyer  
; APPLICANT: and William Charles Ripka  
; TITLE OF INVENTION: BOVINE PANCREATIC TRYPSIN INHIBITOR  
; TITLE OF INVENTION: DERIVED INHIBITORS OF FACTOR Xa  
; NUMBER OF SEQUENCES: 284  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 611 West Sixth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90017

COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44Mb storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/086.630C  
; FILING DATE: July 1, 1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA: including application  
; PRIOR APPLICATION DATA: described below:  
; APPLICATION NUMBER: 07/913,232  
; FILING DATE: July 13, 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 202/210  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 119:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 58 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-086-630C-119

1

Query Match 10.6%; Score 34.5; DB 1; Length 58;  
Best Local Similarity 38.9%; Pred. No. 4.3e+03;  
Matches 7; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Qy 38 FTGDRCONVWASFYKAE 55  
Db 10 YTG-ICRAYIRYFYNK 26

## RESULT 27

US-08-086-630C-121  
; Sequence 121, Application US/08086630C  
; Patent No. 5747449  
; GENERAL INFORMATION:  
; APPLICANT: Ignace Lasters, Marc De Maeyer  
; APPLICANT: and William Charles Ripka  
; TITLE OF INVENTION: BOVINE PANCREATIC TRYPSIN INHIBITOR  
; TITLE OF INVENTION: DERIVED INHIBITORS OF FACTOR Xa  
; NUMBER OF SEQUENCES: 284  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon & Lyon  
STREET: 611 West Sixth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90017  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44Mb storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/086.630C  
FILING DATE: July 1, 1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below:  
APPLICATION NUMBER: 07/913.232  
FILING DATE: July 13, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 202/210  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 121:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 58 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-086-630C-121

Query Match 10.6%; Score 34.5; DB 1; Length 58;  
Best Local Similarity 38.9%; Pred. No. 4.3e+03;  
Matches 7; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Qy 38 FTGDRCONVWASFYKAE 55  
Db 10 YTGP-CRAYITRYFYNK 26

RESULT 28  
US-08-086-630C-145  
Sequence 145, Application US/08086630C  
Patent No. 5747449  
GENERAL INFORMATION:  
APPLICANT: Ignace Lasters, Marc De Maeyer  
APPLICANT: and William Charles Ripka  
TITLE OF INVENTION: BOVINE PANCREATIC TRYPSIN INHIBITOR  
TITLE OF INVENTION: DERIVED INHIBITORS OF FACTOR Xa  
NUMBER OF SEQUENCES: 284  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 611 West Sixth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90017  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44Mb storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/086.630C  
FILING DATE: July 1, 1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below:  
APPLICATION NUMBER: 07/913.232  
FILING DATE: July 13, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 202/210  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 152:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 58 amino acids  
TYPE: amino acid

PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below:  
APPLICATION NUMBER: 07/913.232  
FILING DATE: July 13, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 202/210  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 145:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 58 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-086-630C-145

Query Match 10.6%; Score 34.5; DB 1; Length 58;  
Best Local Similarity 38.9%; Pred. No. 4.3e+03;  
Matches 7; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Qy 38 FTGDRCONVWASFYKAE 55  
Db 10 YTGP-CRAYITRYFYNK 26

RESULT 29  
US-08-086-630C-152  
Sequence 152, Application US/08086630C  
Patent No. 5747449  
GENERAL INFORMATION:  
APPLICANT: Ignace Lasters, Marc De Maeyer  
APPLICANT: and William Charles Ripka  
TITLE OF INVENTION: BOVINE PANCREATIC TRYPSIN INHIBITOR  
TITLE OF INVENTION: DERIVED INHIBITORS OF FACTOR Xa  
NUMBER OF SEQUENCES: 284  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 611 West Sixth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90017  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44Mb storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/086.630C  
FILING DATE: July 1, 1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below:  
APPLICATION NUMBER: 07/913.232  
FILING DATE: July 13, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 202/210  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 152:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 58 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-086-630C-152

Query Match 10.6%; Score 34.5; DB 1; Length 58;  
Best Local Similarity 38.9%; Pred. No. 4.3e+03;  
Matches 7; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Qy 38 FTGDRCONVVMASFYKAE 55  
Db 10 YTGPR-CRAYITRYFYNAK 26

RESULT 30

US-08-086-630C-153  
; Sequence 153, Application US/08086630C  
; Patent No. 5747449  
; GENERAL INFORMATION:  
; APPLICANT: Ignace Lasters, Marc De Maeyer  
; APPLICANT: and William Charles Ripka  
; TITLE OF INVENTION: BOVINE PANCREATIC TRYPSIN INHIBITOR  
; TITLE OF INVENTION: DERIVED INHIBITORS OF FACTOR Xa  
; NUMBER OF SEQUENCES: 284  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 611 West Sixth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90017  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44Mb storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/086,630C  
; FILING DATE: July 1, 1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; PRIOR APPLICATION DATA: including application  
; PRIOR APPLICATION DATA: described below: 1  
; APPLICATION NUMBER: 07/913,232  
; FILING DATE: July 13, 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 202/210  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 153:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 58 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-086-630C-153

Query Match 10.6%; Score 34.5; DB 1; Length 58;  
Best Local Similarity 38.9%; Pred. No. 4.3e+03;  
Matches 7; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Qy 38 FTGDRCONVVMASFYKAE 55  
Db 10 YTGPR-CRAYITRYFYNAK 26

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2005, 08:17:51 ; Search time 247 Seconds  
(without alignments)

90.818 Million cell updates/sec

Title: US-09-980-672-2

Perfect score: 327

Sequence: 1 SHLVKCAKTCVNGGECFM.....TGDRCONVWVSFYKABELY 58

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 6433

Minimum DB seq length: 58

Maximum DB seq length: 58

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

A Geneseqp2004s\*

1: geneseqp1980s\*

2: geneseqp1990s\*

3: geneseqp2000s\*

4: geneseqp2001s\*

5: geneseqp2002s\*

6: geneseqp2003as\*

7: geneseqp2003bs\*

8: geneseqp2004s\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	327	100.0	58	3	AAB12601 Human neu
2	67.5	20.6	58	1	AAP80159 Biosynthe
3	67.5	20.6	58	5	ABB06749 Epidermal
4	65.5	20.0	58	2	AAR27689 eEGF in y
5	62.5	19.1	58	3	AAB08507 Amino aci
6	53	16.2	58	4	AAB06706 Pig trans
7	48	14.7	58	5	ABP07212 Human ORF
8	46	14.1	58	3	AAG03353 Human sec
9	46	14.1	58	5	AAB26505 Human con
10	46	14.1	58	6	ABG72940 Consensus
11	46	14.1	58	6	ABU62263 Consensus
12	46	14.1	58	7	ADP89585 Consensus
13	46	14.1	58	8	ADH60894 Consensus
14	45	13.8	58	4	AAB48112 Human TAN
15	44.5	13.6	58	4	AAB60778 Scorpion
16	43.5	13.3	58	4	AUS0362 Propionib
17	43.5	13.3	58	6	ASW46881 Consensus
18	42.5	13.0	58	3	AG34809 Arabidops
19	42.5	13.0	58	3	AAG08000 Arabidops
20	39	11.9	58	5	AAE14759 Human CCR
21	39	11.9	58	5	AAE14755 Human CCR
22	37.5	11.5	58	4	AAM92968 Human dig
23	37.5	11.5	58	4	AAB60786 Scorpion
24	37.5	11.5	58	4	AAB20064 Human liv
25	37.5	11.5	58	5	ABP40925 Human liv

26	37.5	11.5	58	7	ADJ15043	Human liv
27	37.5	11.5	58	8	ADS85149	Tick (Ixo
28	37	11.3	58	4	ABG09491	Novel hum
29	36	11.0	58	2	AAW40520	Human p75
30	36	11.0	58	5	ABP32785	Human ORF
31	35.5	10.9	58	4	AAAM15170	Peptide #
32	35.5	10.9	58	4	ABB34163	Peptide #
33	35.5	10.9	58	4	AAAM27627	Peptide #
34	35.5	10.9	58	4	ABB28994	Peptide #
35	35.5	10.9	58	4	ABB19606	Protein #
36	35.5	10.9	58	4	AAAM67336	Human bon
37	35.5	10.9	58	4	AAU62273	Propionib
38	35.5	10.9	58	4	AAAM54956	Human bra
39	35.5	10.9	58	4	ABG48999	Human liv
40	35.5	10.9	58	4	ABG21260	Novel hum
41	35.5	10.9	58	5	ABP05278	Human ORF
42	35.5	10.9	58	6	ABM58792	Propionib
43	35	10.7	58	3	AAV65133	Human 5'
44	35	10.7	58	5	ABP08071	Human ORF
45	35	10.7	58	5	AAU10605	Human dis
46	35	10.7	58	6	ABG74277	Human mat
47	34.5	10.6	58	2	AAAR48107	BPTI (IA,
48	34.5	10.6	58	2	AAAR48112	BPTI (IA,
49	34.5	10.6	58	2	AAAR48114	BPTI (IA,
50	34.5	10.6	58	2	AAAR48104	BPTI (IA,
51	34.5	10.6	58	2	AAAR48111	BPTI (IA,
52	34.5	10.6	58	2	AAAR48113	BPTI (IA,
53	34.5	10.6	58	2	AAAR48115	BPTI (IA,
54	34.5	10.6	58	2	AAAR48116	BPTI (IA,
55	34.5	10.6	58	2	AAAR48117	BPTI (IA,
56	34.5	10.6	58	2	AAW79991	Bovine pa
57	34.5	10.6	58	2	AAW80100	Bovine pa
58	34.5	10.6	58	2	AAW80107	Bovine pa
59	34.5	10.6	58	2	AAW80103	Bovine pa
60	34.5	10.6	58	2	AAW80112	Bovine pa
61	34.5	10.6	58	2	AAW80118	Bovine pa
62	34.5	10.6	58	2	AAW80106	Bovine pa
63	34.5	10.6	58	2	AAW80102	Bovine pa
64	34.5	10.6	58	2	AAW79997	Bovine pa
65	34.5	10.6	58	2	AAW80101	Bovine pa
66	34.5	10.6	58	4	AAAM17979	Peptide #
67	34.5	10.6	58	4	ABB37007	Peptide #
68	34.5	10.6	58	4	AAAM30485	Peptide #
69	34.5	10.6	58	4	ABB31778	Peptide #
70	34.5	10.6	58	4	ABB22321	Protein #
71	34.5	10.6	58	4	AAAM70147	Human bon
72	34.5	10.6	58	4	AAAM57734	Human bra
73	34.5	10.6	58	4	ABG51857	Human liv
74	34.5	10.6	58	4	AAAM05614	Peptide #
75	34.5	10.6	58	5	ABG39786	Human pep
76	34	10.4	58	2	AAAR39671	C-termina
77	34	10.4	58	2	AAAR39674	C-termina
78	34	10.4	58	2	AAAR78545	Human col
79	34	10.4	58	2	AAAR81918	Human col
80	34	10.4	58	2	AAAR99204	Human apr
81	34	10.4	58	2	AAW01851	Antileuko
82	34	10.4	58	2	AAW01843	Antileuko
83	34	10.4	58	2	AAW01844	Antileuko
84	34	10.4	58	2	AAW01846	Antileuko
85	34	10.4	58	2	AAW01845	Antileuko
86	34	10.4	58	2	AAW64117	Human Kun
87	34	10.4	58	2	AAW92864	US5880256
88	34	10.4	58	3	AAAB09649	IGFBP-2 I
89	34	10.4	58	3	AAAB09650	IGFBP-2 I
90	34	10.4	58	3	AAAB16806	Bacteriop
91	34	10.4	58	3	AAAB43546	Human can
92	34	10.4	58	3	AAAG45681	Arabidops
93	34	10.4	58	4	AAAE13087	Collagen
94	34	10.4	58	4	AAAB60626	Collagen
95	34	10.4	58	4	AAAU14484	Human nov
96	34	10.4	58	4	AAAU14248	Human nov
97	34	10.4	58	7	ADP42016	Human col
98	34	10.4	58	8	ADF71956	Kallikrei



XX WPI; 2002-257605/30.  
 DR N-PSDB; ABL50267.  
 XX  
 PT Collagen-binding hybrid polypeptide, useful in drug delivery system for  
 PT functional polypeptides, formulating into complex with collagen to give  
 PT functionally modified collagen matrix as biomaterial for tissue  
 PT regeneration.  
 XX  
 PS Example 1; Page 70; 86pp; Japanese.  
 XX  
 CC The present invention describes a hybrid polypeptide comprising a  
 CC collagen-binding domain composed of an amino acid sequence ranging from  
 CC Ala at position 260 to Arg at position 484 of human fibronectin or a  
 CC similar amino acid sequence but with some amino acids deleted,  
 CC substituted, inserted or added, and a functional polypeptide linked  
 CC together. The present invention also describes: (1) a biomaterial  
 CC containing a functional polypeptide-modified collagen obtained by  
 CC formulating a polypeptide originated from the hybrid polypeptide with  
 CC collagen into a complex; (2) a gene encoding the hybrid polypeptide; and  
 CC (3) a transformant containing the gene. The hybrid polypeptide is useful  
 CC in drug delivery system for functional polypeptides, formulating into  
 CC complex with collagen to give functionally modified collagen matrix as  
 CC biomaterial for tissue regeneration. The hybrid polypeptide has superior  
 CC activity and stability in the body for a long period of time, with  
 CC localisability and sustained-releasability. The present sequence  
 CC represents a human epidermal growth factor with an enterokinase  
 CC recognition sequence, which is used in an example from the present  
 CC invention  
 XX  
 SQ Sequence 58 AA;  
 Query Match 20.6%; Score 67.5; DB 5; Length 58;  
 Best Local Similarity 32.4%; Pred. No. 7.8;  
 Matches 11; Conservative 8; Mismatches 12; Indels 3; Gaps 1;  
 Oy 11 FCVNGGCFMWKLSNPSRYLCKPNEFTGDRQC 44  
 Db 18 YCLHDGVCMYTEAL---DKYACNCVGVYIGERCQ 48  
 RESULT 4  
 AAR27689  
 ID AAR27689 standard; protein; 58 AA.  
 AC AAR27689;  
 XX  
 XX 25-MAR-2003 (revised)  
 DT 09-MAR-1993 (first entry)  
 XX  
 XX eEGF in yeast expression vector pTD4-37.  
 DE  
 XX  
 KW Equine; epidermal growth factor; EGF; pTD4-37; Yeast; alpha-factor;  
 KW pre-pro; site-directed mutagenesis; M13mp18; pBE11; pTD4-18;  
 KW lysine-arginine endopeptidase; KEX2; precursor; mouse; human;  
 KW prophylaxis; granulation; horse.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..5  
 FT /note= "Yeast alpha-factor pre-pro sequence"  
 FT Protein 6..58  
 FT /note= "Mature eEGF"  
 XX  
 XX WO9216626-A1.  
 FN  
 XX  
 PD 01-OCT-1992.  
 XX  
 PF 09-MAR-1992; 92WO-GB000416.  
 XX  
 PR 14-MAR-1991; 91GB-00005446.  
 XX

(BRBI-) BRITISH BIO-TECHNOLOGY LTD.  
 Edwards RM, McCullagh KG, Power CAG;  
 WPI; 1992-349211/42.  
 N-PSDB; AAQ28066.  
 Equine epidermal growth factor and variants - useful in healing of wounds  
 and ulcers, esp. for treating tendinitis in horses.  
 Disclosure; Fig 4; 53pp; English.  
 The sequence given shows the equine epidermal growth factor (eEGF) gene  
 within the yeast vector pTD4-37. The N-terminal of the mature eEGF coding  
 residues were linked to the C-terminus coding residues of the yeast alpha  
 -factor pre-pro sequence by site-directed mutagenesis. The resultant  
 fragment was ligated into M13mp18 to give pBE11. This plasmid was cloned  
 and then inserted into the yeast expression vector pTD4-18, replacing the  
 human EGF gene. This gave pTD4-37. This plasmid was expressed and the  
 mature EGF was liberated from the pre-pro sequence by the yeast lysine-  
 arginine endopeptidase KEX2. The mature EGF was then secreted into the  
 culture medium. The mature protein is useful for the treatment or  
 prophylaxis of ulcers and/or wounds, esp. in horses. The use of eEGF in  
 horses may lead to a decrease, or even avoidance, of scarring and  
 formation of granulation tissue scars of the lower limbs. (Updated on 25-  
 MAR-2003 to correct PN field.)  
 SQ Sequence 58 AA;  
 Query Match 20.0%; Score 65.5; DB 2; Length 58;  
 Best Local Similarity 28.6%; Pred. No. 13;  
 Matches 10; Conservative 10; Mismatches 12; Indels 3; Gaps 1;  
 Oy 11 FCVNGGCFMWKLSNPSRYLCKPNEFTGDRQC 45  
 Db 18 YCLHGGKCVYLQVDT---HACNCVGVYIGERCQ 49  
 RESULT 5  
 AAB08507  
 ID AAB08507 standard; protein; 58 AA.  
 AC AAB08507;  
 XX  
 XX 20-DEC-2000 (first entry)  
 DT  
 XX  
 DE Amino acid sequence of modified epidermal growth factor.  
 XX  
 KW Fibronectin; collagen-binding domain; sustained release; gene therapy;  
 KW physiologically active polypeptide; topical retention;  
 KW tissue regeneration; epidermal growth factor.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1..5  
 FT /note= "enterokinase recognition sequence"  
 FT Protein 6..58  
 FT /note= "human epidermal growth factor"  
 FT Misc-difference 16  
 FT /note= "this residue is given as Xaa, as at this position  
 the sequence is off the edge of the page"  
 FT Misc-difference 32  
 FT /note= "this residue is given as Xaa, as at this position  
 the sequence is off the edge of the page"  
 FT Misc-difference 48  
 FT /note= "this residue is given as Xaa, as at this position  
 the sequence is off the edge of the page"  
 XX  
 PN WO200049159-A1.  
 XX  
 PD 24-AUG-2000.

XX PF 21-FEB-2000; 2000WO-JP000964.  
 XX PR 19-FEB-1999; 95JP-00041913.  
 XX PR 01-NOV-1999; 95JP-00311364.  
 XX PA (TERU) TERUMO CORP.  
 XX PI Ishikawa T, Kitajima T;  
 XX WPI; 2000-565375/52.  
 XX DR N-PSDB; AAA64269.  
 XX CC Collagen-binding active polypeptide for use in an agent for enabling  
 PT topical retention or sustained release of a physiologically active  
 PT peptide or physiological activity-imparting agent comprises a fibronectin  
 PT peptide.  
 XX FS Disclosure; Page 117; 135pp; English.  
 XX CC The present sequence represents a modified human epidermal growth factor  
 CC with an enterokinase recognition sequence. The protein is used to  
 CC construct a collagen-binding physiologically active polypeptide. This  
 CC polypeptide comprises a peptide from fibronectin ligated to a  
 CC physiologically active peptide. The polypeptides are used in an agent for  
 CC enabling topical retention or sustained release of a physiologically  
 CC active peptide or physiological activity-imparting agent. They may be  
 CC used in gene therapy and in tissue regeneration  
 XX SQ Sequence 58 AA;  
 Query Match 19.1%; Score 62.5; DB 3; Length 58;  
 Best Local Similarity 30.3%; Pred. No. 27;  
 Matches 10; Conservative 8; Mismatches 12; Indels 3; Gaps 1;  
 Qy 11 FCVNGGECFMVKDLSNPSRYLCKPNEFTGDRG 43  
 Db 18 YCLHDGVCMYIEAL---KKYACNCVGVIGERC 47  
 RESULT 6  
 AA06706  
 ID AAB06706 standard; protein; 58 AA.  
 XX AC AA06706;  
 XX DT 16-OCT-2001 (first entry)  
 XX DE Pig transforming growth factor (TGF) alpha precursor protein.  
 XX KW Pig; TGF; transforming growth factor alpha-like protein;  
 KW nervous system disease; neuropathy; SCID; cancer; gene therapy;  
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease;  
 KW myeloid cell disorder; lymphoid cell disorder; platelet disorder;  
 KW thrombocytopenia; osteoporosis; osteoarthritis; diabetes mellitus;  
 KW bone degenerative disorder; liver fibrosis; autoimmune disorder;  
 KW severe combined immunodeficiency; infection; multiple sclerosis;  
 KW rheumatoid arthritis; periodontal disease; vaccine.  
 XX OS Sus scrofa.  
 XX PN WO200155333-A2.  
 XX PD 02-AUG-2001.  
 XX PF 25-JAN-2001; 2001WO-US002457.  
 XX PR 25-JAN-2000; 2000US-00491404.  
 XX PR 08-AUG-2000; 2000US-00634024.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Boyle BJ, Mize NK, Arterburn MC, Palencia S, Tang YT, Liu C;

PI DRmanac RA, Labat I, Stache-Crane B, Nguyen K, Garcia VE;  
 XX WPI; 2001-488788/53.  
 XX CC Novel transforming growth factor alpha-like polypeptides and  
 PT polynucleotides for diagnosis, treatment of immune, neurological,  
 PT hematopoietic, bone degenerative disorders, bacterial or fungal  
 PT infections and cancer.  
 XX FS Disclosure; Fig 1; 132pp; English.  
 XX CC The invention relates to transforming growth factor (TGF) alpha-like  
 CC polypeptides and polynucleotides. TGF alpha-like sequences exhibit  
 CC cytokine, cell proliferation/differentiation, stem cell growth factor  
 CC activities, immunoglobulin like activity and activin/inhibin related  
 CC activities and hence is useful for re-engineering damaged or diseased  
 CC tissues, transplantation, manufacture of bio-pharmaceuticals and  
 CC development of bio-sensors. It is useful for the treatment of central and  
 CC peripheral nervous system diseases and neuropathies such as Alzheimer's  
 CC disease, Parkinson's disease, Huntington's disease, amyotrophic lateral  
 CC sclerosis and Shy-Drager syndrome; myeloid or lymphoid cell disorders,  
 CC platelet disorders such as thrombocytopenia; regeneration of bone,  
 CC cartilage, tendon, ligament and/or nerve tissue growth, osteoporosis, it  
 CC osteoarthritis, bone degenerative disorders, and periodontal disease. It  
 CC is also useful for gut protection or regeneration; treatment of lung or  
 CC liver fibrosis; reperfusion injury in various tissues; immune  
 CC deficiencies and disorders including severe combined immunodeficiency  
 CC (SCID); bacterial or fungal infections; autoimmune disorders e.g.  
 CC multiple sclerosis, rheumatoid arthritis, diabetes mellitus and for the  
 CC diagnosis, prognosis, treatment of one or more types of cancer. The  
 CC sequences of the invention are used to treat or heal wounds in tissues  
 CC such as skin, cornea and gastrointestinal tract, to promote angiogenesis  
 CC and to monitor the level of expression of TGF-alpha in cancer cells. They  
 CC are also used in gene therapy and as antigens in vaccine compositions to  
 CC raise an immune response. The present sequence is pig transforming growth  
 CC factor (TGF) alpha precursor protein. This sequence is homologous to  
 CC human TGF alpha-like protein of the invention  
 XX SQ Sequence 58 AA;  
 Query Match 16.2%; Score 53; DB 4; Length 58;  
 Best Local Similarity 28.9%; Pred. No. 2.7e+02;  
 Matches 13; Conservative 8; Mismatches 18; Indels 6; Gaps 2;  
 Qy 1 SHLVKC--AEKTCVNGGECFMVKDLSNPSRYLCKPNEFTGDRG 43  
 Db 18 SHFNDCPDSDSHSQFCFHGTCTFLVQE----DKPACVCHSGVVGARC 58  
 RESULT 7  
 ABP07212  
 ID ABP07212 standard; protein; 58 AA.  
 XX AC ABP07212;  
 XX DT 24-JUN-2002 (first entry)  
 XX DE Human ORFX protein sequence SEQ ID NO:14406.  
 XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KW hypertension; hypothyroidism; cholesterol ester storage disease;  
 KW immune deficiency; immune disorder; infectious disease;  
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KW myasthenia gravis.  
 XX OS Homo sapiens.  
 XX PN WO200192523-A2.  
 XX PD 06-DEC-2001.



```

XX 29-MAY-2001; 2001WO-US010836.
XX
XX 30-MAY-2000; 2000US-0206132P.
XX
XX 29-AUG-2000; 2000US-0228716P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach MD;
XX
XX WPI; 2002-106308/14.
XX
XX N-PSDB; ABN22964.
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cardiovascular disease, neurodegenerative,
XX hyperproliferative disorders and autoimmune disorders.
XX
XX Disclosure; SEQ ID NO 14406; 1037pp; English.
XX
XX The present invention describes substantially purified human proteins
XX (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
XX in the specification). ABN15762 to ABN27252 encode the human ORFX
XX proteins given in ABP0010 to ABP11500. ORFX proteins are useful for
XX treating or preventing a pathology associated with an ORFX-associated
XX disorder in humans, and in the manufacture of a medicament for treating a
XX syndrome associated with ORFX-associated disorder. ORFX polynucleotide
XX sequences can be used in gene therapy. ORFX sequences can be used in the
XX treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
XX psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
XX osteoarthritis, neurodegenerative disorders, disorders related to organ
XX transplantation, cardiovascular diseases, diabetes mellitus, systemic
XX lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
XX storage disease, various immune deficiencies and disorders, infectious
XX diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
XX arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
XX disease and autoimmune inflammatory eye disease. ORFX proteins are also
XX useful for treating burns, incisions, ulcers, for treating osteoporosis,
XX bone degenerative disorders, or periodontal disease, and for gut
XX protection or regeneration and treatment of lung or liver fibrosis,
XX reperfusion injury in various tissues and conditions resulting from
XX systemic cytokine damage. N.B. The sequence data for this patent did not
XX form part of the printed specification, but was obtained in electronic
XX format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 58 AA;
XX
XX Query Match 14.7%; Score 48; DB 5; Length 58;
XX Best Local Similarity 34.2%; Pred. No. 9.4e+02;
XX Matches 13; Conservative 4; Mismatches 15; Indels 6; Gaps 3;
XX
XX QY 12 CVNGGCEFMVKDLSNPSRYLCKPNE-FTGDRCCQNYVM 48
XX | : || | | | | | | | | | | | | | | | | | | |
XX 22 CAHGTC---HDLVNGFR--CDCACTGVGEGTHCEREVL 54
XX
XX RESULT 8
XX AAG03353
XX ID AAG03353 standard; protein; 58 AA.
XX
XX AC AAG03353;
XX
XX XX AAG03353;
XX
XX DT 06-OCT-2000 (first entry)
XX
XX DE Human secreted protein, SEQ ID NO: 7434.
XX
XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping.
XX
XX OS Homo sapiens.
XX
XX PN EP1033401-A2.
XX
XX XX 06-SEP-2000.
XX

```

```

XX 21-FEB-2000; 2000EP-00200610.
XX
XX 26-FEB-1999; 99US-0122487P.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI; 2000-500381/45.
XX
XX N-PSDB; AAC03359.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
XX Claim 13; SEQ ID NO 7434; 71pp + Sequence Listing; English.
XX
XX The present sequence is a polypeptide encoded by one of a large number of
XX 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
XX prepared from total human RNAs or polyA+ RNAs derived from 30 different
XX tissues. EST sequences usually correspond mainly to the 3' untranslated
XX region (UTR) of the mRNA because they are often obtained from oligo-dT
XX primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
XX sequences derived from the 5' ends of mRNAs and even in those cases where
XX longer cDNA sequences have been obtained, the full 5' UTR is rarely
XX included. 5' ESTs are derived from mRNAs with intact 5' ends and can
XX therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
XX are also used in diagnostic, forensic, gene therapy and chromosome
XX mapping procedures. They are used to obtain upstream regulatory sequences
XX and to design expression and secretion vectors
XX
XX Sequence 58 AA;
XX
XX Query Match 14.1%; Score 46; DB 3; Length 58;
XX Best Local Similarity 36.4%; Pred. No. 1.5e+03;
XX Matches 12; Conservative 4; Mismatches 9; Indels 8; Gaps 3;
XX
XX QY 6 CAEKTCVNGGCEP--MYKDLNPSRYLCKCPN 36
XX | : | | | | | | | | | | | | | | | | | | |
XX 10 CRVRVAC-----CFANMVFVXHNPKSFLC-CPH 36
XX
XX RESULT 9
XX AAE26505
XX ID AAE26505 standard; peptide; 58 AA.
XX
XX AC AAE26505;
XX
XX DT 13-DEC-2002 (first entry)
XX
XX DE Human consensus EGF-R repeat peptide.
XX
XX KW Antibody; epidermal growth factor; EGF repeat; brain tumour;
XX nervous disorder; ulcer; leukaemia; human.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 39..41
XX /label= Unknown
XX /note= "Xaa can be any amino acid"
XX FT Misc-difference 45..46
XX /label= Unknown
XX /note= "Xaa can be any amino acid"
XX
XX US6392019-B1.
XX
XX 21-MAY-2002.
XX
XX 28-JUL-1999; 99US-00363316.
XX
XX 22-NOV-1997; 97US-00968800.
XX

```

```

PR 12-FEB-1999; 99US-00249697.
XX (FORD/) FORD J.
PA (YEUN/) YEUNG G.
XX
PI Ford J, Yeung G;
XX WPI; 2002-424836/45.
DR
XX Novel antibody specific for an epidermal growth factor repeat-containing
PT polypeptide, useful for the diagnosis of brain tumors, ulcers, leukemias,
PT and nervous disorders.
XX
PS Disclosure; Col 77-80; 92pp; English.
XX
CC The invention relates to an antibody specific for a 537 residue epidermal
CC growth factor (EGF) repeat-containing polypeptide sequence. The invention
CC is used for detecting the presence of EGF repeat containing polypeptides
CC in a sample, in the diagnosis of brain tumors, nervous disorders,
CC ulcers, and leukemias. The present sequence is human consensus EGF-R
CC repeat peptide
XX
SQ Sequence 58 AA;
Query Match 14.1%; Score 46; DB 5; Length 58;
Best Local Similarity 28.2%; Pred. No. 1.5e+03;
Matches 11; Conservative 8; Mismatches 12; Indels 8; Gaps 3;
QY 11 FCVNGGECFMVKDLSNPSRYLCKCPNEF---TGDR--CQ 44
DB 13 YCLHDGVCMTIEAL---DKYACNCVVGVIYXXGGERXXCQ 48
RESULT 10
ABG72940
ID ABG72940 standard; protein; 58 AA.
XX
AC ABG72940;
XX
DT 02-APR-2003 (first entry)
XX
DE Consensus epidermal growth factor (EGF)-repeat motif.
XX
KW EGF; epidermal growth factor; cancer; lung cancer; brain cancer;
KW prostate cancer; breast cancer; skin cancer; lymphoma cancer;
KW sarcoma cancer; colon cancer; tumorigenicity; tumour site reduction;
KW cell proliferation inhibition; vaccine; antisense gene therapy;
KW EGF-repeat motif.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 39 /label= OTHER
FT /note= "OTHER= Any amino acid"
FT Misc-difference 40 /label= OTHER
FT /note= "OTHER= Any amino acid"
FT Misc-difference 41 /label= OTHER
FT /note= "OTHER= Any amino acid"
FT Misc-difference 45 /label= OTHER
FT /note= "OTHER= Any amino acid"
FT Misc-difference 46 /label= OTHER
FT /note= "OTHER= Any amino acid"
FT Misc-difference 47 /label= OTHER
FT /note= "OTHER= Any amino acid"
US2002132250-A1.
XX
PD 19-SEP-2002.
XX
PP 15-OCT-2001; 2001US-00981649.

```

```

XX 28-JUL-1999; 99US-00363316.
PR 13-OCT-2000; 2000US-00687860.
XX
PA (FORD/) FORD J E.
PA (YEUN/) YEUNG G.
PA (ZHOU/) ZHOU H.
XX
PI Ford JE, Yeung G, Zhou H;
XX WPI; 2003-174078/17.
DR
XX Detecting cancerous cells expressing polynucleotides/polypeptides in
PT samples, by contacting samples with labeled polynucleotides complementary
PT to polynucleotide or an antibody against the polypeptide and detecting
PT complex formed.
XX
XX Example 4; Fig 1; 78pp; English.
XX
CC The invention describes a method of detecting a cancerous cell expressing
CC a polynucleotide (I) or a polypeptide (II) in a biological sample,
CC involving contacting the sample with a labelled polynucleotide
CC complementary to (I) or an antibody or its fragment that specifically
CC binds to (II), for a period sufficient to form a complex and detecting
CC the complex, so that if a complex is detected, the cell is detected. The
CC method is useful for detecting cancerous cell in a biological sample such
CC as tissue, cell, blood, serum, lymphatic fluid, urine, and cerebrospinal
CC fluid. The cancerous cell is from lung, brain, prostate, breast, skin,
CC lymphoma, sarcoma and colon. Preferably the cancer cell is A549 cell, MCF
CC -7 cell or SK-N-Mc cell. PC1 and PC2 are useful for inhibiting EGF16
CC activity, inhibiting tumorigenicity, reducing tumour sites and inhibiting
CC proliferation of a cancer cell. This is the amino acid sequence of
CC consensus motif for epidermal growth factor (EGF) repeat motifs
XX
SQ Sequence 58 AA;
Query Match 14.1%; Score 46; DB 6; Length 58;
Best Local Similarity 28.2%; Pred. No. 1.5e+03;
Matches 11; Conservative 8; Mismatches 12; Indels 8; Gaps 3;
QY 11 FCVNGGECFMVKDLSNPSRYLCKCPNEF---TGDR--CQ 44
DB 13 YCLHDGVCMTIEAL---DKYACNCVVGVIYXXGGERXXCQ 48
RESULT 11
ABU62263
ID ABU62263 standard; protein; 58 AA.
XX
AC ABU62263;
XX
DT 01-SEP-2003 (first entry)
XX
DE Consensus epidermal growth factor repeat.
XX
KW Epidermal growth factor motif protein; EGF16; cytostatic;
KW neuroprotective; antibacterial; antiparasitic; antileptic;
KW antiinfertility; EGF-agonist; EGF-Antagonist; cell growth; cancer;
KW neurodegenerative disorder; leukaemia; brain tumour; lung tumour;
KW breast tumour; gastrointestinal tumour; skin tumour; prostate tumour;
KW carcinoma; parasite; biorhythm; fertility; metabolism; catabolism;
KW anabolism; consensus; epidermal growth factor repeat.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 39 /label= OTHER
FT /note= "OTHER= any amino acid"
FT Misc-difference 40 /label= OTHER
FT /note= "OTHER= any amino acid"
FT Misc-difference 41

```





XX PD 28-DEC-2000.  
XX PF 21-JUN-2000; 2000WO-US017049.  
XX PR 22-JUN-1999; 99US-0140227P.  
XX FA (DUPO ) DU PONT DE NEMOURS & CO E I.  
XX PI Herrmann R, Lee J, Wong JF;  
XX WPI; 2001-071394/08.  
XX DR New polynucleotides encoding scorpion venom potassium-channel agonist  
PT proteins for production e.g. of insect-tolerant transgenic plants for  
PT controlling insect pest damage and parasitic worm infections.  
XX PS Claim 10; Fig 1; 50pp; English.  
XX CC The present invention relates to scorpion toxins. The invention may be  
CC used for the creation of transgenic plants which express K-channel  
CC modifiers, useful as a means for controlling insect pests by producing  
CC insect-tolerant plants. In the prevention and/or treatment of insect pest  
CC damage and parasitic worm infections in animals and humans, the invention  
CC may also find use in creating specific new pesticides and antihelmintic  
CC drugs that are also non-toxic to humans, pets and livestock  
XX SQ Sequence 58 AA;  
Query Match 13.6%; Score 44.5; DB 4; Length 58;  
Best Local Similarity 38.7%; Pred. No. 2.2e+03;  
Matches 12; Conservative 1; Mismatches 7; Indels 11; Gaps 2;  
Qy 4 VKCAEKTFCVNGGCEFMVKDLNPSRYLCKC 34  
Db 37 VKCAMGF-NSGKINSK-----CKC 56  
RESULT 16  
AAU50362  
ID AAU50362 standard; protein; 58 AA.  
XX AC AAU50362;  
XX DT 13-FEB-2002 (first entry)  
XX DE Propionibacterium acnes immunogenic protein #11258.  
XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.  
XX OS Propionibacterium acnes.  
XX PN WO200181581-A2.  
XX PD 01-NOV-2001.  
XX PF 20-APR-2001; 2001WO-US012865.  
XX PR 21-APR-2000; 2000US-0199047P.  
XX PR 02-JUN-2000; 2000US-0208841P.  
XX PR 07-JUL-2000; 2000US-0216747P.  
XX PA (CORI-) CORIXA CORP.  
XX SK Skey YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX WPI; 2001-616774/71.  
DR N-PSDB; AAS59548.  
XX

PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
XX treating acne vulgaris.  
XX PS Example 1; SEQ ID NO 11557; 1069pp; English.  
XX CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 58 AA;  
Query Match 13.3%; Score 43.5; DB 4; Length 58;  
Best Local Similarity 26.5%; Pred. No. 2.8e+03;  
Matches 9; Conservative 7; Mismatches 9; Indels 9; Gaps 1;  
Qy 28 SRYLC-----KCPNEFTGRCQNYVMASFY 52  
Db 23 SRYLWRLVGRSRACPSQFPGRSLQFLGALW 56  
RESULT 17  
ABM46881  
ID ABM46881 standard; protein; 58 AA.  
XX AC ABM46881;  
XX DT 20-OCT-2003 (first entry)  
XX DE Propionibacterium acnes predicted ORF-encoded polypeptide #11557.  
XX KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
KW immunostimulant; immune response; vaccine.  
XX OS Propionibacterium acnes.  
XX PN WO2003033515-A1.  
XX PD 24-APR-2003.  
XX PF 11-OCT-2002; 2002WO-US032727.  
XX PR 15-OCT-2001; 2001US-00978825.  
XX PA (CORI-) CORIXA CORP.  
XX PI Mitcham JL, Skey YAW, Persing DH, Bhatia A, Maisonneuve JL;  
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
PI Barth B, Vallie-Douglas J;  
XX WPI; 2003-381789/36.  
DR N-PSDB; ACF64477.  
XX New Propionibacterium acnes polypeptides and polynucleotides encoding the  
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
PT or for stimulating an immune response specific for a P. acnes protein.  
PT



```
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157553P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158002P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159291P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.

PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 13.0%; Score 42.5; DB 3; Length 58;
Best Local Similarity 28.6%; Pred. No. 3.6e+03;
Matches 10; Conservative 4; Mismatches 8; Indels 13; Gaps 2;

QY 12 CVNGGECFMVKDLNPSRYLCKPNE-PTGDRCON 45
Db 27 CLNSQCAKA-----CPSEGFSGRCSS 49

RESULT 19
AAG08000
ID AAG08000 standard; protein; 58 AA.
XX
AC AAG08000;
XX
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 5367.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
```

PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135623P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140699P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142053P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144003P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.

PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145132P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.



```

PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 26-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 13.0%; Score 42.5; DB 3; Length 58;
Best Local Similarity 28.6%; Pred. No. 3.6e+03;
Matches 10; Conservative 4; Mismatches 8; Indels 13; Gaps 2;

QY 12 CVNGGECFWMKDLNPSRYLCKPNE-FTGDRCON 45
Db 27 CLNSOSCAK-----CPSEGFSGGRCS 49

RESULT 20
AAE14759
ID AAE14759 standard; peptide; 58 AA.
XX
AC AAE14759;
XX
DT 04-OCT-2002 (first entry)
XX
DE Human CCR5 chemokine receptor C-terminal mutant fragment (aa 295-352).
XX
KW Chemokine receptor CCR5; P2 protein; HIV infection; AIDS;
KW human immunodeficiency virus; gene therapy; human; mutant; mutein.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 58
FT /note= "Wild-type Leu replaced by Pro"
XX
PN EPI207202-A1.
XX
PD 22-MAY-2002.
XX
PF 12-NOV-2001; 2001EP-00126500.
XX
PR 16-NOV-2000; 2000EP-00125052.
XX
PA (MOEL/) MOELLING K.
XX
PI Moelling K, Schwenecker M;
XX
DR WPI; 2002-437464/47.
XX
CC Nucleic acid encoding a protein which interacts with the carboxy terminus
CC of the chemokine receptor CCR5 isolated from a human B cell cDNA library
CC is useful to detect, treat and prevent HIV infection.
XX
PS Example 2; Fig 2A; 28pp; English.
XX
CC The invention relates to nucleic acid encoding a protein which interacts
CC with the carboxy terminus of the chemokine receptor CCR5 or of other
CC chemokine receptor family members. Molecules of the invention are used to
CC detect, prevent and treat HIV infection and progression to AIDS. Any CCR5
CC derived peptides, or compounds derived from the protein of the invention
CC (e.g. antagonist or mutants) that disrupt binding of CCR5 to the protein
CC of the chemokine receptor CCR5 isolated from a human B cell cDNA library
CC is useful to detect, treat and prevent HIV infection.
XX
PS Example 2; Fig 2A; 28pp; English.
XX
CC The invention relates to nucleic acid encoding a protein which interacts
CC with the carboxy terminus of the chemokine receptor CCR5 or of other
CC chemokine receptor family members. Molecules of the invention are used to
CC detect, prevent and treat HIV infection and progression to AIDS. Any CCR5
CC derived peptides, or compounds derived from the protein of the invention

```

```

CC (e.g. antagonist or mutants) that disrupt binding of CCR5 to the protein
CC of the invention would block HIV infection. The present sequence is human
CC chemokine receptor CCR5 C-terminal mutant fragment which is incapable of
CC interacting with the protein of the invention
XX
SQ Sequence 58 AA;

Query Match 11.9%; Score 39; DB 5; Length 58;
Best Local Similarity 37.5%; Pred. No. 8.5e+03;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 38 FTGDRCONVWASFYK 53
Db 5 FVGEKFRNYLLVFFQK 20

RESULT 21
AAE14755
ID AAE14755 standard; peptide; 58 AA.
XX
AC AAE14755;
XX
DT 04-OCT-2002 (first entry)
XX
DE Human CCR5 chemokine receptor C-terminal fragment (residues 295-352).
XX
KW Chemokine receptor CCR5; P2 protein; HIV infection;
KW human immunodeficiency virus; AIDS; gene therapy; human.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Binding-site 58
FT /note= "Essential for binding to human protein P2
FT (AAE14754)"
XX
PN EPI207202-A1.
XX
PD 22-MAY-2002.
XX
PF 12-NOV-2001; 2001EP-00126500.
XX
PR 16-NOV-2000; 2000EP-00125052.
XX
PA (MOEL/) MOELLING K.
XX
PI Moelling K, Schwenecker M;
XX
DR WPI; 2002-437464/47.
XX
CC Nucleic acid encoding a protein which interacts with the carboxy terminus
CC of the chemokine receptor CCR5 isolated from a human B cell cDNA library
CC is useful to detect, treat and prevent HIV infection.
XX
PS Example 2; Fig 2A; 28pp; English.
XX
CC The invention relates to nucleic acid encoding a protein which interacts
CC with the carboxy terminus of the chemokine receptor CCR5 or of other
CC chemokine receptor family members. Molecules of the invention are used to
CC detect, prevent and treat HIV infection and progression to AIDS. Any CCR5
CC derived peptides, or compounds derived from the protein of the invention
CC (e.g. antagonist or mutants) that disrupt binding of CCR5 to the protein
CC of the invention would block HIV infection. The present sequence is human
CC chemokine receptor CCR5 C-terminal fragment
XX
SQ Sequence 58 AA;

Query Match 11.9%; Score 39; DB 5; Length 58;
Best Local Similarity 37.5%; Pred. No. 8.5e+03;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 38 FTGDRCONVWASFYK 53
| |:::|:::|

```

Db 5 FVGEKFRNYLLVFFOK 20

RESULT 22  
AAM92968  
ID AAM92968 standard; protein; 58 AA.  
AC  
XX AAM92968;  
DT 05-NOV-2001 (first entry)  
XX Human digestive system antigen SEQ ID NO: 2317.  
DE Human; digestive system antigen; gene therapy; cancer; appendicitis;  
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
KW digestive system disorder; Meckel's diverticulum.  
XX  
OS Homo sapiens.  
XX  
PN WO200155314-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001324.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205151P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225211P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249255P.  
PR 17-NOV-2000; 2000US-0249297P.

```

PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251989P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX PA
XX (HUMA-) HUMAN GENOME SCI INC.
XX PI
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-502630/55.
XX DR N-PSDB; AAK88741.
XX PT
XX Polynucleotides encoding digestive system antigens, useful for
XX diagnosing, treating, preventing and/or prognosing disorders of the
XX digestive system, particularly cancer and cancer metastases.
XX Claim 11; SEQ ID NO 2317; 986pp; English.
XX CC
XX The present invention provides the protein and coding sequences of a
XX number of human digestive system antigens. These can be used in the
XX diagnosis, treatment and prevention of digestive system disorders,
XX including cancer, Meckel's diverticulum, bacterial or parasitic
XX infections, appendicitis, Hirschsprung's disease, chronic colitis or
XX ulcerative colitis. The present sequence is a digestive system antigen of
XX the invention
XX SQ
XX Sequence 58 AA;
XX Query Match 11.5%; Score 37.5; DB 4; Length 58;
XX Best Local Similarity 77.8%; Pred. No. 1.2e+04;
XX Matches 7; Conservative 1; Mismatches 0; Indels 1; Gaps 1;
XX
XX Qy 30 YLCKCPNEF 38
XX Db 13 HLC-CPNEF 20
XX
XX RESULT 23
XX AAB60786
XX ID AAB60786 standard; protein; 58 AA.
XX AC AAB60786;
XX XX
XX DT 28-MAR-2001 (first entry)
XX DE Scorpion potassium channel blocking toxin 15-1 protein #2.
XX XX
XX Scorpion; toxin; K-channel; potassium; insect; pesticide.
XX OS
XX Hottentotta judaica.
XX XX
XX PN WO200078958-A2.
XX XX
XX PD 28-DEC-2000.
XX XX
XX PF 21-JUN-2000; 2000WO-US017049.
XX XX
XX PR 22-JUN-1999; 99US-0140227P.
XX XX
XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX XX
XX PI Herrmann R, Lee J, Wong JF;

```

```

XX WPI; 2001-071394/08.
XX New polynucleotides encoding scorpion venom potassium-channel agonist
XX PT proteins for production e.g. of insect-tolerant transgenic plants for
XX PT controlling insect pest damage and parasitic worm infections.
XX PS Claim 10; Fig 7; 50pp; English.
XX CC
XX The present invention relates to scorpion toxins. The invention may be
XX used for the creation of transgenic plants which express K-channel
XX modifiers, useful as a means for controlling insect pests by producing
XX insect-tolerant plants. In the prevention and/or treatment of insect pest
XX damage and parasitic worm infections in animals and humans, the invention
XX CC may also find use in creating specific new pesticides and antihelmintic
XX CC drugs that are also non-toxic to humans, pets and livestock
XX SQ
XX Sequence 58 AA;
XX Query Match 11.5%; Score 37.5; DB 4; Length 58;
XX Best Local Similarity 27.8%; Pred. No. 1.2e+04;
XX Matches 10; Conservative 10; Mismatches 9; Indels 7; Gaps 3;
XX
XX Qy 12 CVNGGECFMV-KDLSNPSRYLCKPCNFTGDRCONY 46
XX Db 29 CISSQECWIACKKVT--GRPQKCKNK----QCRCY 58
XX
XX RESULT 24
XX AAU20064
XX ID AAU20064 standard; protein; 58 AA.
XX XX
XX AC AAU20064;
XX XX
XX DT 04-DEC-2001 (first entry)
XX DE Human liver associated polypeptide #95.
XX XX
XX Liver associated protein; human; mouse; rabbit; goat; horse; cat; dog;
XX KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
XX KW antirheumatic; antiproliferative; cytostatic; cardiac; neuroprotective;
XX KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
XX KW ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;
XX KW hyperproliferative disorder; breast; liver; cardiovascular disorder;
XX KW cerebrovascular disorder; nervous system disorder; bacterial infection;
XX KW fungal infection; viral infection; ocular disorder; endocrine disorder;
XX KW gastrointestinal disorder; renal disorder; respiratory disorder;
XX KW wound healing; skin aging; organ transplantation; tissue regeneration;
XX KW anti-infertility.
XX XX
XX OS Homo sapiens.
XX XX
XX PN WO200155355-A1.
XX XX
XX PD 02-AUG-2001.
XX XX
XX PF 17-JAN-2001; 2001WO-US001351.
XX XX
XX 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180628P.
XX PR 24-FEB-2000; 2000US-0184664P.
XX PR 02-MAR-2000; 2000US-0186350P.
XX PR 16-MAR-2000; 2000US-0189874P.
XX PR 17-MAR-2000; 2000US-0190076P.
XX PR 18-APR-2000; 2000US-0198123P.
XX PR 19-MAY-2000; 2000US-0205515P.
XX PR 07-JUN-2000; 2000US-0209467P.
XX PR 28-JUN-2000; 2000US-0214886P.
XX PR 30-JUN-2000; 2000US-0215135P.
XX PR 07-JUL-2000; 2000US-0216647P.
XX PR 07-JUL-2000; 2000US-0216880P.
XX PR 11-JUL-2000; 2000US-0217487P.
XX PR 11-JUL-2000; 2000US-0217496P.

```



CC viruses and fungi, ocular disorders such as corneal infection, endocrine  
 CC disorders such as premature labour and infertility, gastrointestinal  
 CC disorders such as Crohn's disease, renal disorders such as  
 CC glomerulonephritis and respiratory disorders such as asthma and pleurisy.  
 CC The polypeptides can also be used to aid wound healing, to prevent skin  
 CC aging due to sunburn, to maintain organs before transplantation, to  
 CC regenerate tissues and in chemotaxis. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

Query Match 11.5%; Score 37.5; DB 4; Length 58;  
 Best Local Similarity 77.8%; Pred. No. 1.2e+04;  
 Matches 7; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Oy 30 YLCKCPNEF 38  
 :||| |||||  
 Db 13 HLC-CPNEF 20

## RESULT 25

ABP40925

ID ABP40925 standard; protein; 58 AA.

XX AC ABP40925;

XX DT 24-JUL-2002 (first entry)

XX DE Human liver antigen HLBW64, SEQ ID NO:251.

XX KW Human; liver antigen; liver disorder; hepatic disorder; infection;  
 KW hepatitis; viral; parasitic; bacterial; fungal; inflammatory condition;  
 KW cirrhosis; granulomatous hepatitis; toxin damage; drug damage;  
 KW autoimmune disease; Wilson's disease; primary biliary cirrhosis;  
 KW neoplastic disorder; cancer; tumour; portal hypertension;  
 KW gastrointestinal disorder; hepatitis; drug screening; gene therapy;  
 KW chromosome mapping; forensic analysis; antibody preparation;  
 KW hepatotropic; cytostatic; antineoplastic; virucide; antibacterial;  
 KW fungicide; parasiticide; antidote; immunosuppressive.

XX OS Homo sapiens.

XX PN US2002042096-A1.

XX PD 11-APR-2002.

XX PF 17-JAN-2001; 2001US-00764887.

XX PR 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 28-JUN-2000; 2000US-0214886P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 22-AUG-2000; 2000US-0226868P.

PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.

PR 01-SEP-2000; 2000US-0229344P.

PR 01-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 13-OCT-2000; 2000US-0237040P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.

(ROSE/) ROSEN C A.

(RUBE/) RUBEN S M.

(BARA/) BARASH S C.

Rosen CA, Ruben SM, Barash SC;

WPI; 2002-381944/41.

N-PSDB; ABN90130.

New nucleic acid encoding human liver antigens, useful for diagnosis, treatment and prevention of e.g. hepatitis and hepatic cancer, also related polypeptides and antibodies.

Claim 11; SEQ ID NO 251; 181pp; English.

The invention relates to 145 novel human liver antigens (ABP40811-ABP40975) and to cDNAs encoding them (ABN90036-ABN90180), and also encompasses polypeptides 90% identical and polynucleotides 95% identical to the sequences of the invention. The invention additionally relates to recombinant vectors and host cells comprising human liver antigen polynucleotides, antibodies against human liver antigens, and the use of liver antigen polynucleotides and polypeptides in diagnosing, treating, prognosing or preventing various disorders of the liver. Such conditions include viral infections (e.g., cytomegalovirus, Epstein-Barr virus, hepatitis A virus, hepatitis B virus and hepatitis C virus), parasitic infections (e.g., Clonorchis sinensis, Echinococcus granulosus and Entamoeba histolytica), and also bacterial and fungal infections. Other disorders that may be treated include inflammatory conditions (e.g., cirrhosis and granulomatous hepatitis), damage caused by drugs or toxins, autoimmune diseases (e.g., Wilson's disease, primary biliary cirrhosis), neoplastic disorders (e.g., adenomas, haemangiomas and hepatocellular carcinoma), portal hypertension, or gastrointestinal disorders (e.g., peptic ulcers, gastritis and peritoneal diseases). Liver antigen polypeptides and polynucleotides may also be used in screening for compounds which modulate liver antigen expression or activity. The polynucleotides may further be used for gene therapy, chromosome mapping, in the identification of individuals and in forensic analysis, and the polypeptides may be used as molecular weight markers or to prepare antibodies useful in disease diagnosis, drug targeting and phenotyping. The present sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the USPTO at seqdata.uspto.gov/sequence/

Sequence 58 AA;

59

Query Match 11.5%; Score 37.5; DB 5; Length 58;  
Best Local Similarity 77.8%; Pred. No. 1.2e+04;  
Matches 7; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 30 YLCKCPNEF 38  
Db 13 HLC-CPNEF 20

RESULT 26

ADJ15043  
ID ADJ15043 standard; protein; 58 AA.

XX AC

ADJ15043;

DT 20-MAY-2004 (first entry)

XX DE Human liver-related protein - SEQ ID 251.

XX KW liver; virucide; fungicide; antibacterial; antiparasitic; hepatotropic;  
KW antiinflammatory; cytostatic; litholytic; antirheumatic; antiarthritic;  
KW neuroprotective; antidiabetic; anticoagulant; thrombolytic;  
KW antiarteriosclerotic; cardiac; haemostatic; antiarrhythmic;  
KW ophthalmological; antiarteriosclerotic; vasotropic; osteopathic;  
KW neotropic; antiparkinsonian; anticonvulsant; neuroleptic; vasotropic;  
KW cytostatic; gynaecological; viral; fungal; bacterial;  
KW parasitic infection; cirrhosis; Wilson's disease;  
KW gastrointestinal disorder; pancreatic; gallbladder; immune; blood;  
KW hyperproliferative; cardiovascular; respiratory; musculoskeletal system;  
KW neurological; endocrine; reproductive system; developmental; inherited;  
KW human.

XX OS Homo sapiens.

PN US2003077602-A1.

XX PD

XX 24-APR-2003.

XX PF 14-FEB-2002; 2002US-00073961.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226688P.

PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.

PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 PR 17-JAN-2001; 2001US-00764887.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM, Barash SC;  
 XX  
 DR WPI; 2003-765398/72.  
 DR N-PSDB; ADJ14897.  
 XX  
 PT New liver related polypeptide, useful for diagnosis, treatment and/or  
 PT prevention of liver, gastrointestinal, pancreatic, immune, blood related,  
 PT endocrine, reproductive, hyperproliferative or reproductive disorders.  
 XX  
 PS Claim 11; SEQ ID NO 251; 181pp; English.  
 XX  
 CC The invention relates to a novel isolated, liver related polypeptide. The  
 CC polypeptide of the invention demonstrates virucide, fungicide,  
 CC antibacterial, antiparasitic, hepatotropic, antiinflammatory, cytostatic,  
 CC litholytic, antirheumatic, antiarthritic, neuroprotective, antidiabetic,  
 CC anticoagulant, thrombolytic, antiarteriosclerotic, cardiant, haemostatic,  
 CC antiarrhythmic, ophthalmological, antiarteriosclerotic, vasotropic,  
 CC osteopathic, nootropic, antiparkinsonian, anticonvulsant, neuroleptic,  
 CC vasotropic, cytostatic and gynaecological activities. The polypeptides  
 CC and polynucleotides of the invention may be useful for diagnosis,  
 CC detection, treatment and/or prevention of disorders of the liver such as  
 CC viral, fungal, bacterial or parasitic infections, cirrhosis, Wilson's  
 CC disease, gastrointestinal disorders, pancreatic disorders, gallbladder  
 CC diseases, immune intestinal disorders, blood related disorders, hyperproliferative  
 CC disorders, cardiovascular disorders, respiratory disorders,  
 CC musculoskeletal system disorders, neurological diseases, endocrine  
 CC disorders, reproductive system disorders or developmental and inherited  
 CC disorders. The current sequence is that of the human liver-related  
 CC protein of the invention. The current sequence is not shown within the  
 CC specification per se but was obtained electronically from the USPTO web-  
 CC site.

Query Match 11.5%; Score 37.5; DB 7; Length 58;  
 Best Local Similarity 77.8%; Pred. NO. 1.2e+04;  
 Matches 7; Conservative 1; Mismatches 0; Indels 1; Gaps 1;  
 QY 30 YLCKPNEF 38  
 DB 13 HLC-CFNEF 20

RESULT 27  
 ADS85149  
 ID ADS85149 standard; protein; 58 AA.  
 XX  
 AC ADS85149;  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX  
 DE Tick (Ixodes scapularis) protein - L2.  
 XX  
 KW tick; tick-borne pathogen; tick-borne disease; L2.  
 XX  
 OS Ixodes scapularis.  
 XX  
 PN WO2004019883-A2.  
 XX  
 PD 11-MAR-2004.  
 XX  
 PF 29-AUG-2003; 2003WO-US027071.  
 XX  
 PR 30-AUG-2002; 2002US-0408218P.  
 XX  
 PA (UYVA ) UNIV YALE  
 PA (LTWO-) L2 DIAGNOSTICS LLC.  
 XX  
 PI Fikrig E, Kantor F, Narasimhan S, Cappello M, Koski RA;  
 PI Ledizet M;  
 XX  
 DR WPI; 2004-239112/22.  
 XX  
 PT New tick polypeptide, useful for preparing a composition for preventing  
 PT infection by a tick-borne pathogen or a tick-borne disease.  
 XX  
 PS Example 8; Page 40; 87pp; English.  
 XX  
 CC The invention comprise the amino acid and coding sequences of tick  
 CC proteins. The DNA and protein sequences of the invention are useful for  
 CC preparing a composition for preventing infection by a tick-borne pathogen  
 CC or tick-borne disease. The present amino acid sequence represents a tick  
 CC protein of the invention.  
 XX  
 SQ Sequence 58 AA;  
 Query Match 11.5%; Score 37.5; DB 8; Length 58;  
 Best Local Similarity 32.1%; Pred. NO. 1.2e+04;  
 Matches 9; Conservative 4; Mismatches 8; Indels 7; Gaps 1;  
 QY 8 EKTFCVNG-----GECFVVKDLSNPS 28  
 DB 28 ERCEFYNGDEGLCQNGECLTRDSGVFN 55

RESULT 28  
 ABG09491  
 ID ABG09491 standard; protein; 58 AA.  
 XX  
 AC ABG09491;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #9482.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX

30-MAR-2001; 2001WO-US008631.  
 31-MAR-2000; 2000US-00540217.  
 23-AUG-2000; 2000US-00649167.  
 (HYSE-) HYSEQ INC.  
 Drmanac RT, Liu C, Tang YT;  
 WPI, 2001-639362/73.  
 N-FSDB; AAS73678.  
 New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.  
 Claim 20; SEQ ID NO 39850; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activities. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 58 AA;  
 Query Match 11.3%; Score 37; DB 4; Length 58;  
 Best Local Similarity 60.0%; Pred. No. 1.4e+04;  
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 5 KCAEKTFCVN 14  
 Db 41 ECEEKRLCVN 50

RESULT 29  
 AAW40520  
 ID AAW40520 standard; protein; 58 AA.

AC AAW40520;  
 XX 15-JUL-1998 (first entry)  
 DT Human p75 neurotrophin binding domain fragment.

Nerve growth factor; NGF; optical structure; active conformation;  
 variable basis Monte Carlo stimulated annealing; VBMC; ligand;  
 neurotrophin; agonist; inhibitor; TrkA; TrkB; TrkC; NT3; p75NTR;  
 brain-derived neurotrophic factor; human.

Homo sapiens.  
 Key Location/Qualifiers  
 Domain 1..41  
 Disulfide-bond 1..17  
 /label= Second cysteine-rich domain

Disulfide-bond 20..33  
 Disulfide-bond 23..41  
 Domain 42..58  
 /label= Third cysteine-rich domain  
 Disulfide-bond 43..56  
 WO9806048-A2.  
 12-FEB-1998.  
 31-JUL-1997; 97WO-CA000539.  
 31-JUL-1996; 96GB-00016105.  
 (TOOH) UNIV QUEBENS KINGSTON.  
 Shamovsky IL, Ross GM, Riopelle RJ, Weaver DF;  
 WPI, 1998-145797/13.  
 Identifying optimal molecular structures by variable basis Monte Carlo method - particularly to determine biologically active conformation(s) of neurotrophin domains involved in receptor binding and subsequent evolution of active ligands.  
 Disclosure; Fig 17; 170pp; English.  
 Protein fragments AAW0504-W40523 are used in a variable basis Monte Carlo (VEMC) stimulated annealing method for identifying an optimal molecular structure. The method is used to identify the biologically active conformations of peptide domains of ligands, particularly neurotrophins (NT), that bind to a receptor. Once such conformations have been identified, small molecules that either inhibit NT-binding or act as agonists of the native ligand can be developed. Specifically the method is applied to (a) nerve growth factor/TrkA; (b) brain-derived neurotrophic factor, NT3 or 4/TrkB; (c) NT3/TrkC and (d) the receptor p75NTR. The preferred ligand comprises elements with proper spatial occupancy, relative atomic positions, bond type and charge for defining a 3-dimensional configuration able to bind the 3 binding areas. TrkA includes a leucine-rich motif (LRM), amino acids 93-117, with 5 binding areas: (A) Phe105TA (hydrophobic interaction); (B) Phe111TA, Phe113TA and Thr114TA (hydrophobic interaction); (C) Asp109TA and His112TA (ionic); (D) Lys100TA (ionic) and (E) Asn95TA-Ile98TA (multiple parallel beta - strand type hydrogen bonds)

Sequence 58 AA;

Query Match 11.0%; Score 36; DB 2; Length 58;  
 Best Local Similarity 28.3%; Pred. No. 1.8e+04;  
 Matches 15; Conservative 8; Mismatches 18; Indels 12; Gaps 4;

Qy 1 SHLVKCAEKTFCVNGGECFVWKDLSNP----SRYLKCK-----PNEFTGDRQC 44  
 Db 8 SDVVSATEP--CKPCTECVGLQNSAPCVREADDVACRCAYGYQDETGTG-RCE 57

RESULT 30  
 ABP32785  
 ID ABP32785 standard; protein; 58 AA.

AC ABP32785;  
 XX 08-JUL-2002 (first entry)  
 DT Human ORF1758 protein, SEQ ID NO:3516.

Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
 disease monitoring; cytokine; cell proliferation; cell differentiation;  
 immune modulation; haematopoiesis regulation; tissue growth;  
 angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;  
 thrombolytic; tumour inhibition; bodily characteristic; fertility;  
 behaviour; cancer; proliferative disorder; neurological disorder;  
 cardiovascular disease; immune system disorder; organ transplantation;



KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;  
 KW vasotropic; antipsoriatic; antidiabetic; cytostatic; neutropic;  
 KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;  
 KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
 KW dermatological; analgesic; virucide; antibacterial; fungicide.

OS Homo sapiens.

XX WO200190366-A2.

XX PD 29-NOV-2001.

XX PF 24-MAY-2001; 2001WO-US017076.

XX PR 24-MAY-2000; 2000US-0206690P.

XX PA (CURA-) CURAGEN CORP.

XX PI Leach MD, Shinkets RA;

XX WPI; 2002-106200/14.

DR N-PSDB; ABN76811.

XX Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and disorders related to organ  
 PT transplantation.

PS Claim 10; Page 1139; 2508pp; English.

XX Sequences ABP1028-ABP3561 represent 4534 novel human proteins  
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
 CC ABN79587 represent cDNAs encoding them. The invention also encompasses  
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
 CC specific for ORFX proteins, methods for detecting ORFX polynucleotides and  
 CC polypeptides, methods of screening for modulators of ORFX expression or  
 CC activity, and methods of screening individuals for a predisposition to an  
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 CC range of biological activities, such as cytokine, cell proliferation,  
 CC cell differentiation, immune modulation, haematopoiesis regulation,  
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,  
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
 CC and antiinfective activity, and may also be involved in the determination  
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
 CC nucleic acids and antibodies may be used in the treatment of cancers,  
 CC other proliferative disorders such as psoriasis and benign tumours,  
 CC neurological disorders such as epilepsy and Alzheimer's disease,  
 CC cardiovascular diseases, immune system disorders, disorders related to  
 CC organ transplantation, disorders of tissue growth and regeneration,  
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 CC storage disease, and infectious diseases caused by viral, bacterial,  
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
 CC source of primers and probes, in the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the  
 CC diagnosis, treatment and monitoring of ORFX-associated diseases

XX Sequence 58 AA;

Query Match 11.0%; Score 36; DB 5; Length 58;

Best Local Similarity 46.2%; Pred. No. 1.8e+04;

Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 23 DLSNPSRYLCKCP 35

Db 29 EFSPPSRWCHFP 41

Search completed: April 14, 2005, 08:44:04  
 Job time : 256 secs

This Page Blank (uspio)